Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники
Направление подготовки 09.04.03 Прикладная информатика
Отделение информационных технологий

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Информационная технология оценки показателей качества жизни пациентов

УДК 004:304.3-047.43:616-052

Студент

Группа	ФИО	Подпись	Дата
8КМ61	Былина Татьяна Андреевна		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Марухина Ольга Владимировна	к.т.н.		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Старикова Екатерина Васильевна	к.ф.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ОКД	Авдеева Ирина Ивановна	-		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП 09.04.03 Прикладная информатика	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Марухина Ольга Владимировна	к.т.н.		

Томск – 2018 г.

Запланированные результаты обучения по программе

Код результата	Результат обучения (выпускник должен быть готов)
Профессиональные компетенции	
Р1	Применять базовые и специальные знания в области современных информационно-коммуникационных технологий для решения междисциплинарных инженерных задач.
Р2	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретацию полученных данных в области информатизации и автоматизации прикладных процессов и создания, внедрения, эксплуатации и управления информационными системами в прикладных областях.
Р3	Внедрять, сопровождать и эксплуатировать современные информационные системы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья и безопасности труда, выполнять требования по защите окружающей среды.
Р4	Активно владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты инновационной инженерной деятельности.
Р5	Владеть и применять методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе глобальных компьютерных сетей.
Р6	Эффективно работать индивидуально, в качестве члена и руководителя группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации.
Р7	Самостоятельно учиться и непрерывно повышать квалификацию в течение всего периода профессиональной деятельности.
Профиль «Системы корпоративного управления»	
Р8	Применять глубокие профессиональные знания основ построения информационных технологий и систем, достаточные для решения научных и профессиональных задач производства. Знать современные проблемы и методы

Код результата	Результат обучения (выпускник должен быть готов)
	прикладной информатики и научно-технического развития информационных технологий.
Р9	Ставить и решать задачи комплексного анализа, связанные с информатизацией и автоматизацией прикладных процессов; созданием, внедрением, эксплуатацией и управлением информационными системами в прикладных областях, с использованием базовых и специальных знаний, современных аналитических методов и моделей.
Р10	Организовывать работы по моделированию прикладных информационных систем и реинжинирингу прикладных и информационных процессов предприятия и организации. Управлять проектами по информатизации прикладных задач и созданию информационных систем предприятий и организаций.

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники
Направление подготовки 09.04.03 Прикладная информатика
Отделение информационных технологий

УТВЕРЖДАЮ:
Руководитель ООП
_____ _____ Марухина О.В.
(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

магистерской диссертации

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8КМ61	Былиной Татьяне Андреевне

Тема работы:

Информационная технология оценки показателей качества жизни пациентов	
Утверждена приказом директора (дата, номер)	20.04.2018г. №2796/с

Срок сдачи студентом выполненной работы:	

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	Результаты тестирования пациентов по методикам оценки показателей качества жизни
Перечень подлежащих исследованию, проектированию и разработке вопросов	1. Постановка задачи 2. Выбор методов решения задачи 3. Решение поставленной задачи 4. Описание полученных результатов
Перечень графического материала	Схема алгоритма обработки данных Диаграммы размаха Диаграммы визуализации пропущенных значений Графический результат кластеризации

Консультанты по разделам выпускной квалификационной работы	
Раздел	**Консультант**
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Старикова Екатерина Васильевна, к.ф.н., доцент ОСГН
Социальная ответственность	Авдеева Ирина Ивановна, ассистент ОКД
Обязательное приложение на английском языке	Краснова Татьяна Ивановна, старший преподаватель ОИЯ

Названия разделов, которые должны быть написаны на русском и иностранном языках:
Оценка качества жизни
Методы обработки данных
Разработка технологии оценки качества жизни
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение
Социальная ответственность
Methods of Initial Data Analysis

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	05.02.2018

Задание выдал руководитель:

Должность	**ФИО**	**Ученая степень, звание**	**Подпись**	**Дата**
Доцент ОИТ ИШИТР	Марухина О. В.	к. т. н.		

Задание принял к исполнению студент:

Группа	**ФИО**	**Подпись**	**Дата**
8КМ61	Былина Т.А.		

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники
Направление подготовки 09.04.03. Прикладная информатика
Уровень образования Магистратура
Отделение Информационных технологий
Период выполнения Весенний семестр 2018 учебного года
Форма представления работы:

магистерская диссертация

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
05.02.2018	Получение задания на ВКР	
01.03.2018	Получение задания по финансовому менеджменту	
01.03.2018	Получение задания по социальной ответственности	
05.03.2018	Глава 1. Оценка качества жизни	
26.03.2018	Глава 2. Методы обработки данных	
16.04.2018	Глава 3. Разработка технологии оценки качества жизни	
30.04.2018	Глава 4. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	
14.05.2018	Глава 5. Социальная ответственность	
21.05.2018	Оформление работы.	
29.05.2018	Сдача выполненной работы.	

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Марухина Ольга Владимировна	к. т. н.		

СОГЛАСОВАНО:

Руководитель ООП 09.04.03 Прикладная информатика	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Марухина Ольга Владимировна	к. т. н.		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8КМ61	Былиной Татьяне Андреевне

Школа	Инженерная школа информационных технологий и робототехники	**Отделение**	Информационных технологий
Уровень образования	Магистратура	**Направление/ специальность**	09.04.03 Прикладная информатика

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:	
1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Информационная технология разработана на основе анализа исходных психологических данных, которые представляют собой совокупность результатов тестирования пациентов по различным методикам оценки качества жизни.
2. Нормы и нормативы расходования ресурсов	
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Оценка коммерческого и инновационного потенциала НТИ	Оценка конкурентоспособности разработки, анализ перспективности исследования
2. Планирование и формирование бюджета научных исследований	Планирование этапов разработки технологии, определение трудоемкости, построение диаграммы Ганта, формирование бюджета НТИ.
3. Определение ресурсной, финансовой, экономической эффективности	Сравнительный анализ интегральных показателей эффективности
Перечень графического материала	
1. Сегментирование рынка *2.* Оценка конкурентоспособности технических решений *3.* Оценочная карта QuaD, Матрица SWOT *4.* График проведения и бюджет НТИ *5.* Оценка ресурсной, финансовой и экономической эффективности НТИ	

Дата выдачи задания для раздела по линейному графику	01.03.2018

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Старикова Екатерина Васильевна	к.ф.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8КМ61	Былина Татьяна Андреевна		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8КМ61	Былиной Татьяне Андреевне

Школа	Инженерная школа информационных технологий и робототехники	**Отделение**	Информационных технологий
Уровень образования	Магистратура	**Направление/специальность**	09.04.03 Прикладная информатика

Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Рабочая зона оборудована 10 рабочими местами, каждое из которых включает в себя персональный компьютер, расположенный на персональном столе и офисный стул. Площадь аудитории составляет 36 м^2, в аудитории имеется 2 оборудованных жалюзи окна размером 1*1,3 м., выходящих на южную сторону, поэтому естественное освещение можно использовать в дневное время, в остальное время используется равномерное искусственное освещение. Стены и потолок имеют светлые оттенки, пол застелен линолеумом светло-серого оттенка.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Производственная безопасность 1.1. Анализ выявленных вредных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности: – физико-химическая природа вредности, её связь с разрабатываемой темой; – действие фактора на организм человека; – приведение допустимых норм с необходимой размерностью (со ссылкой на соответствующий нормативно-технический документ); – предлагаемые средства защиты; – (сначала коллективной защиты, затем – индивидуальные защитные средства). 1.2. Анализ выявленных опасных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности: – механические опасности (источники, средства защиты; – термические опасности (источники, средства защиты); – электробезопасность (в т.ч. статическое электричество, молниезащита – источники, средства защиты); – пожаровзрывобезопасность (причины, профилактические мероприятия, первичные средства пожаротушения).	Анализ выявленных вредных факторов: – повышенный уровень электромагнитных излучений; – повышенная или пониженная влажность воздуха; – повышенная или пониженная температура воздуха; – повышенный уровень шума; – недостаточная освещенность рабочего места. Анализ психофизических факторов: – эмоциональные перегрузки; – умственное перенапряжение; – монотонность труда; – перенапряжение зрения. Анализ выявленных опасных факторов: – опасность поражения электрическим током; – вероятность возникновения короткого замыкания; – статическое электричество.
2. Экологическая безопасность: – защита селитебной зоны; – анализ воздействия объекта на атмосферу (выбросы);	Анализ воздействия объекта на литосферу: утилизация отходов, связанные с выходом из строя ПК, люминесцентных ламп и др.

– анализ воздействия объекта на гидросферу (сбросы); – анализ воздействия объекта на литосферу (отходы); – разработать решения по обеспечению экологической безопасности со ссылками на НТД по охране окружающей среды.	
3. Безопасность в чрезвычайных ситуациях: – перечень возможных ЧС при разработке и эксплуатации проектируемого решения; – выбор наиболее типичной ЧС; – разработка превентивных мер по предупреждению ЧС; – разработка действий в результате возникшей ЧС и мер по ликвидации её последствий.	Возможные чрезвычайные ситуации на объекте: – Пожар (наиболее вероятен); – Землетрясение.
4. Правовые и организационные вопросы обеспечения безопасности: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	– СанПиН 2.2.2/2.4.1340 – 03 – организация рабочих мест с электронно-вычислительными машинами. – ГОСТ 12.2.032-78 ССБТ – организация рабочего места при выполнении работ сидя. – Закон Томской области от 9 июля 2003 года № 83-ОЗ Об охране труда в Томской области (с изменениями на 4 июля 2014 года). – "Трудовой кодекс Российской Федерации" от 30.12.2001 N 197-ФЗ (ред. от 05.02.2018)

Дата выдачи задания для раздела по линейному графику	01.03.2018

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Авдеева Ирина Ивановна			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8КМ61	Былина Татьяна Андреевна		

РЕФЕРАТ

Выпускная квалификационная работа содержит 105 страниц, 21 рисунок, 25 таблиц, 42 источника, 1 приложение.

Ключевые слова: оценка качества жизни, Data Mining, выбросы, восстановление данных, кластерный анализ, факторный анализ.

Объект исследования – качество жизни пациентов.

Цель исследования – разработка технологии оценки показателей качества жизни пациентов (на основе данных тестирования по методикам оценки качества жизни).

В процессе исследования были изучены основные аспекты оценки качества жизни пациентов, методики ее оценки. Во введении приведено обоснование актуальности исследования, выявлена проблема, описаны цель, задачи, предмет и объект исследования. В первой главе описано понятие качества жизни и используемые методики его оценки. Вторая глава содержит описание методов обработки данных. Третья глава посвящена описанию процесса разработки новой информационной технологии оценки показателей качества жизни пациентов. В четвертой главе была обоснована экономическая эффективность проводимого исследования, а пятая – содержит описание социальной ответственности.

Оглавление

Введение

Понятие качества жизни было введено в середине XX века, на тот момент стало необходимо решать проблемы благосостояния населения. В настоящее время это понятие внедрилось в различные области человеческой деятельности, такие как социология, экономика, политика медицина и другие.

Актуальность проведения оценки качества жизни в медицинских исследованиях заключается в том, что эффективность лечения зависит не только от применения различных медико-биологические методов, но и от общего состояния пациента и его моральной готовности к лечению, которые, в свою очередь, зависят от качества жизни.

В настоящее время разработано множество методик для оценки качества жизни. Психологическая служба, проводящая сбор данных столкнулась с проблемой некачественного заполнения листов с тестами. Это связано с тем, что пациенты могут иметь плохое самочувствие, им предлагают большое количество тестов, и ответы на последние они дают с чувством усталости, что отражается на полученных выводах.

Целью работы является разработка технологии оценки показателей качества жизни пациентов (на основе данных тестирования по методикам оценки качества жизни).

Для достижения поставленной цели были определены следующие задачи:

- провести обзор литературы и интернет-источников для определения методов и инструментов исследования;
- исследовать данные на наличие выбросов;
- удалить возможные выбросы и, по возможности, очистить данные от них;
- восстановить однородность исследуемых данных;

- снизить размерность данных за счет выявления латентных переменных;
- в размерности новых переменных провести кластеризацию данных (выявить группы пациентов, сходных по результатам тестирования);
- определить набор тестов, которые будут применяться комплексно для оценки качества жизни пациентов; провести сравнительную оценку временных затрат на тестирования до и после исследования;
- провести финансовый анализ ресурсоэффективности и ресурсосбережения;
- рассмотреть основные аспекты социальной ответственности.

Объектом данного исследования является качество жизни пациентов. Предмет исследования – массив данных тестирования пациентов с применением различных методик оценки качества жизни, среди которых тест SF-36 health status survey, цветовой тест Люшера, шкала базисных убеждений Янов-Бульман WAS, тест жизнестойкости Мадди (в адаптации Леонтьева) и опросник самоорганизации деятельности.

Научная новизна работы заключается в разработке алгоритма определения набора методик для оценки качества жизни пациентов, лежащего в основе разработанной информационной технологии.

Глава 1. Оценка качества жизни

1.1 Описание предметной области

Понятие «качество жизни» в современном обществе широко распространилось и используется довольно часто. Согласно определению, представленному в современном экономическом словаре Райзберга Б.А., качество жизни – оценка некоторого набора условий и характеристик жизни человека, основанная на его собственной степени удовлетворённости этими условиями и характеристиками [1].

Качество жизни может зависеть от различных объективных и субъективных факторов, таких как состояние здоровья, условия окружающей среды, психологический и социальный статусы, ожидаемая продолжительность жизни, коммуникации в обществе, социальное окружение, питание, организованность досуга, удовлетворение культурных и духовных потребностей, бытовой комфорт, профессиональная самореализация, психотип и другие.[2]

Как считают социологи Финансового университета при Правительстве РФ, высокое качество жизни человека подразумевает[3-4]:

- отсутствие угроз для жизни и здоровья людей и качественное медицинское обслуживание, обеспечивающие достаточную продолжительность здоровой жизни;
- обеспечение доступа к материальным благам и необходимого объема потребления товаров и услуг;
- благополучие семьи;
- отсутствие конфликтов в социуме и возможных угроз уровню благополучия;
- возможность познания и развития – открыты доступ к образованию, культурным и духовным ценностям, формированию общего представления об окружающем мире и личности;

- принятие во внимание собственного мнения каждого человека при решении проблем, возникающих в обществе, причастность в создании общепринятых правил поведения человека;
- возможность принимать участие в культурной и общественной жизни, чувство социальной нужности и принадлежности;
- информационная доступность – возможность получать информацию об изменениях, происходящих в различных сферах деятельности человека;
- недолгая продолжительность рабочего дня, которая должна оставлять место для досуга и саморазвития, комфортные условия труда.

В данной работе понятие качества жизни употребляется в контексте медицинских исследований, поэтому для данной области понятие несколько сужается. Как представлено в руководстве по исследованию качества жизни в медицине Новика А.А, качество жизни — это интегральная характеристика физического, психологического, социального и эмоционального состояния пациента, оцениваемая исходя из его субъективного восприятия.[5]

Выделяют два основных аспекта концепции качества жизни. С одной стороны, необходимо принимать во внимание различные стороны жизни пациента, которые могут оказывать прямое или косвенное воздействие на состояние здоровья человека, – физическая, социальная, психологическая, экономическая, культурная и духовная сферы.[5]

Качество жизни исследуют в разных разделах медицины: кардиология, трансплантология, хирургия, психиатрия, онкология, геронтология, неврология, паллиативная медицина и другие. Эти исследования направлены на:

- изучение новых лекарств и методов лечения;
- стандартизацию методов лечения;
- создание моделей течения болезней;
- исследование эффективности лечения;
- экономическое обоснование методов лечения.

1.2 Описание исходных данных

Исходные данные представляют собой двумерный массив, составленный по результатам тестирования пациентов по пяти методикам оценки качества жизни. В опросе приняли участие 663 человека, общее число показателей по всем методикам – 51.

Методики, используемые для тестирования:

1) Sf-36 health status survey:

Это широко используемый в странах Европы и США опросник, который используется для оценки качества жизни пациентаов. Методика описывает общее благополучие и уровень удовлетворенности сторонами жизнедеятельности, зависящими от состояния здоровья.

SF-36 Health Status Survey включает в себя 36 вопросов, которые сгруппированы в восемь шкал:

- Физическое функционирование (Physical Functioning PF) – отражает степень, в которой здоровье позволяет выполнять физические нагрузки (перенос тяжестей, ходьба, самообслуживание, подъем по лестнице и т.п.);
- Ролевая деятельность (Role-Physical RP) – действие физического состояния на ежедневную ролевую деятельность (деятельность, связанная с работой, будничные нагрузки);
- Телесная боль (Bodily Pain BP) – уровень проявления боли и влияние боли на занятия повседневными делами, в том числе деятельность по дому и за его пределами;
- Общее здоровье (General Health GH) – субъективная оценка пациентом состояния своего здоровья и эффективности лечения;
- Жизнеспособность (Vitality VT) – оценка присутствия или отсутствия сил для занятия повседневной деятельностью;
- Социальное функционирование (Social Functioning SF) – степень, в которой общее состояние человека (эмоциональное и физическое) лимитирует его коммуникабельную активность в обществе;

- Эмоциональное состояние (Role-Emotional RE) – подобно ролевой деятельности, оценивает, насколько эмоциональное состояние пациента ограничивает его в занятии повседневной работой;
- Психическое здоровье (Mental Health MH) – субъективная оценка психического здоровья, которая может включать в себя уровень тревожности и стрессов, вероятность возникновения или наличие депрессии, совокупность положительных эмоций и другие.[6]

Все эти шкалы объединяют в 2 большие группы – физический компонент здоровья (физическое функционирование, ролевая деятельность, телесная боль, общее здоровье) и психологический компонент здоровья (жизнеспособность, социальное функционирование, эмоциональное состояние и психическое здоровье).

По каждой шкале сумма показателей лежит в интервале от 0 до 100. Результаты интерпретируются таким образом, что чем выше балл, тем лучше интегральный показатель по выбранной шкале.

2) Цветовой тест Люшера:

Основан на том, что каждый человек воспринимает цвета объективно и универсально, но свои предпочтения субъективно отдает тому или иному цвету. Это позволяет оценить субъективное внутреннее состояния с помощью цветового теста. Тест был адаптирован Л.Н. Собчик, которая разработала систему критериев для оценки тестов Люшера.[7]

Тест включает в себя 4 основных и 4 дополнительных цвета. К основным цветам относятся синий, сине-зеленый, оранжево-красный и светло-желтый. Фиолетовый, коричневый, черный и серый являются дополнительными цветами.[8]

Описание значений цветов представлено в таблице 1.

Таблица 1. Значение цветов по Люшеру

Группа	Цвет	Описание
Основные	Синий	Спокойствие, удовлетворенность
	Сине-зеленый	Уверенность, настойчивость, упрямство

	Оранжево-красный	Сила воли, агрессивность, возбуждение
	Светло-желтый	Активность, стремление к общению, экспансивность, веселость
Дополнительные	Фиолетовый	Тревожность, стресс, переживание страха, огорчения
	Коричневый	
	Черный	
	Серый	

В интерпретации теста большое значение имеет позиция цвета в списке. В результате интерпретации теста можно рассчитать показатель работоспособности, значения уровня стресса и психоэмоциональную напряженность.

3) Шкала базисных убеждений WAS:

Опросник был разработан в 1989 году американским психологом Ронни Янов-Бульман, а в 2007 году адаптирован О.Кравцовой. Он основывается на концепции познания базовых убеждений индивида. Главным предназначением методики является анализ пациентов, переживших клиническую травму и, вероятнее всего, находившихся в состоянии депрессии. Методика позволяет выделить наиболее проблемные области когнитивной сферы, коррекция которых возможна в процессе психотерапии.[9]

Опросник включает в себя 32 утверждения, с которыми пациент должен высказать степень согласия (по шкале от 1 до 6). Утверждения разбиты на 8 шкал: благосклонность мира (benevolence of world BW), доброта людей (benevolence of people BP), справедливость мира (justice J), контролируемость мира (control C), случайность как принцип распределения происходящих событий (randomness R), ценность собственного «Я» (self-worth SW), степень самоконтроля (self-control SC) и степень удачи или везения (luckiness L).

Утверждения могут входить в шкалу как в прямом, так и в переносном значении. Если утверждение входит в обратном значении, то

балл по нему высчитывается путем вычитания указанного балла из семи. В результате баллы по шкалам находят как среднее арифметическое. Таким образом, положительным считается результат выше 3,5 баллов.[9]

4) Тест жизнестойкости Мадди (в адаптации Леонтьева):

Тест был разработан в 1984 году американским психологом Сальваторе Мадди. Адаптацию на русски язык выполнили Д.А.Леонтьев и Е.И. Рассказова. Адаптированный тест состоит из 45 утверждений, которые оцениваются по трёхбалльной шкале (от 0 до 3). Все утверждения разбивают на 3 основные группы: вовлеченность, контроль и принятие риска.

Вовлеченность (commitment) определяется как «убежденность в том, что вовлеченность в происходящее дает максимальный шанс найти нечто стоящее и интересное для личности».[10] Человек с развитым компонентом вовлеченности получает удовольствие от собственной деятельности. В противоположность этому, отсутствие подобной убежденности порождает чувство отвергнутости, ощущение себя «вне» жизни. «Если вы чувствуете уверенность в себе и в том, что мир великодушен, вам присуща вовлеченность». [10]

Контроль (control) представляет собой убежденность в том, что борьба позволяет повлиять на результат происходящего, пусть даже это влияние не абсолютно и успех не гарантирован. Противоположность этому — ощущение собственной беспомощности. Человек с сильно развитым компонентом контроля ощущает, что сам выбирает собственную деятельность, свой путь.

Принятие риска (challenge) — убежденность человека в том, что все то, что с ним случается, способствует его развитию за счет знаний, извлекаемых из опыта, — неважно, позитивного или негативного. Человек, рассматривающий жизнь как способ приобретения опыта, готов действовать в отсутствие надежных гарантий успеха, на свой страх и риск, считая стремление к простому комфорту и безопасности обедняющим жизнь личности. В основе принятия риска лежит идея развития через активное усвоение знаний из опыта и последующее их использование.[10]

Жизнестойкость рассчитывается как сумма трех показателей.

Нормативные значения [10] представлены в таблице 2.

Таблица 2. Нормативные показатели теста жизнестойкости Мадди

Нормы	**Жизнестойкость**	**Вовлечённость**	**Контроль**	**Принятие риска**
Среднее значение	80,72	37,64	29,17	13,91
Стандартное отклонение	18,53	8,08	8,43	4,39

5) Опросник самоорганизации деятельности:

Это методика, образованная при переводе и расширенной адаптации англоязычного опросника структуры времени (Time Structure Questionnaire). Адаптация на русском языке выполнена Мандриковой Е.Ю.

Первоначально был осуществлен прямой перевод с английского на русский язык оригинального варианта TSQ, состоящего из 26 вопросов. В процессе адаптации вопросительная форма была заменена на утвердительную и к каждой группе пунктов, входящей в один из пяти факторов, были добавлены по десять сформулированных авторами суждений, описывающих заявленные в названии факторов конструкты. Таким образом, к двадцати переведенным пунктам добавились еще 50 пунктов.

Опросник включает в себя 6 шкал:

- Шкала «Планомерность» измеряет степень вовлеченности субъекта в тактическое ежедневное планирование по определенным принципам;
- Шкала «Целеустремленность» измеряет способность субъекта сконцентрироваться на цели;
- Шкала «Настойчивость» измеряет склонность субъекта к приложению волевых усилий для завершения начатого дела и упорядочения активности;
- Шкала «Фиксация» измеряет склонность субъекта к фиксации на заранее запланированной структуре организации событий во времени,

его привязанность к четкому расписанию, ригидность в отношении планирования;

- Шкала «Самоорганизация» измеряет склонность субъекта к использованию внешних средств организации деятельности;
- Шкала «Ориентация на настоящее» измеряет временную ориентацию на настоящее.[11]

Баллы по группе суммируются. Нормативные значения [11] представлены в таблице 3.

Таблица 3. Нормативные показатели опросника самоорганизации деятельности

Шкала	**Мужчины**		**Женщины**	
	Среднее значение	**Стандартное отклонение**	**Среднее значение**	**Стандартное отклонение**
Планомерность	19,03	4,61	17,41	5,43
Целеустремленность	32,96	4,79	32,48	7,13
Настойчивость	19,57	5,49	22,19	6,21
Фиксация	19,19	4,75	18,47	5,45
Самоорганизация	9,99	5,00	9,49	4,14
Ориентация на настоящее	8,51	1,86	8,27	3,19
Общий показатель	109,24	15,13	108,30	19,02

Глава 2. Методы обработки данных

2.1 Обзор методов Data Mining

Технология Data Mining представляет собой совокупность различных методов, позволяющих осуществлять самостоятельный поиск нетривиальных зависимостей и закономерностей между данными и формировать предположения, которые помогают лицу, принимающему решение, в изучении поставленной задачи.

К основным методам Data Mining можно отнести следующие:

Технология Data Mining включает в себя достаточно большое количество методов. В этой статье будут рассмотрены наиболее распространенные из них:

- корреляционно-регрессионный анализ используется для того, чтобы производить поиск связей между двумя случайными величинами. В процессе анализа может быть выявлено наличие прямой или обратной связи или её отсутствие;
- дерево решений может также именоваться деревом принятия решений, регрессионным деревом или деревом классификации. Представляет собой иерархическую структуру, построение которой осуществляется по набору определенных правил, представленных в виде конструкций «Если …, то …». Промежуточные узлы и ребра отражают правила, а конечные интерпретируют «корзины», в которые помещаются классифицируемые данные;
- иерархическая кластеризация. Суть метода заключается в пошаговом объединении малых кластеров в более крупные или же, наоборот, в разделении больших кластеров на более мелкие в зависимости от условий поставленной задачи;
- неиерархическая кластеризация. Метод имеет итеративную природу. Разбиение на кластеры происходит до тех пор, пока не будет выполнено правило остановки;

- искусственная нейронная сеть представляет собой модель организации данных и процессов, интерпретирующую работу нервных клеток в организме. Отдельные узлы сети достаточно просты, но, находясь в определенной, четко заданной взаимосвязи, способны решать достаточно сложные задачи;
- эволюционное программирование основано на генетических алгоритмах, которые представляют собой эвристические алгоритмы поиска, производящие подбор и сочетание необходимых данных, применяя механизмы по аналогии с естественным отбором;
- метод опорных векторов. Задачей данного метода является переход из области начальных векторов в новое пространство, которое имеет большую размерность, чем исходное, и поиск в этом пространстве разделяющей гиперплоскости, имеющей больший зазор;
- байесовская сеть представляет собой вероятностную модель, представленную в виде графа, в котором вершины содержат случайные переменные, а ребра соответствуют вероятностным взаимосвязям между ними по Байесу;
- методы ближайшего соседа и k-ближайшего соседа основаны на оценке сходства рассматриваемых объектов. Первый метод полагается на единственный ближайший сходный объект обучающей выборки, второй же менее «доверчив» и требует поиска сходств с k-ближайшими похожими объектами;
- линейная регрессия отображается в виде регрессионной модели, которая описывает линейную взаимосвязь зависимой переменной от одной или нескольких независимых переменных;
- метод поиска ассоциативных правил позволяет находить ассоциативные правила, на основании которых будет приниматься решение. Задача поиска правил включает в себя нахождение часто встречающихся

наборов элементов и генерацию правил из ранее найденных наборов с определенной пороговой достоверностью;

- метод ограниченного перебора используется для нахождения логических взаимосвязей и закономерностей в многомерных массивах данных. Алгоритмы данного метода находят частоты комбинаций простых логических событий в подгруппах данных, на основании анализа которых исследуется полезность комбинации для решения поставленной задачи;
- факторный анализ используется для поиска взаимосвязей между значениями переменных используется факторный анализ. В его основе лежит предположение о том, что все переменные в имеющейся выборке зависят от меньшего числа латентных переменных и случайной ошибки.

Основные задачи, решаемые методами Data Mining

Все выше описанные методы предназначены для решения определенных задач. Все задачи условно могут быть разделены на 6 больших классов:

1. Классификация (стратификация) – нахождение у рассматриваемых объектов специфических признаков, которые определяют их отношение к одному из заранее заданных классов.
2. Кластеризация – это несколько более трудная задача, решаемая методами Data Mining. Для этой задачи классы заранее неизвестны, их необходимо сформировать. В остальном идеи классификации сохраняются.
3. Ассоциация – выявление закономерностей среди взаимосвязанных событий. Основывается на рассмотрении одновременно произошедших событий и выявляется зависимость между произошедшими явлениями.

4. Последовательность (поиск последовательных шаблонов) – нахождение закономерностей среди взаимосвязанных по времени событий.
5. Регрессия и прогнозирование – поиск зависимости выходных данных от входных переменных и предсказание новых результатов на основе выявленных зависимостей.
6. Визуализация – графическое отображение анализируемых данных. Большие объемы сырых данных отображаются в виде наглядных таблиц, схем, диаграмм, графов и т.д.[12]

2.2 Методы анализа исходных данных

2.2.1 Анализ выбросов

Выбросами называют показатели, которые значительно отличаются от остальных в пределах рассматриваемой выборки.[13]

Можно выделить несколько причин, в результате которых образуются выбросы:

- ошибка измерений;
- ошибка ввода данных;
- ошибка интерпретации данных;
- особенности природы данных.

Для получения более точных результатов анализа, выбросы необходимо удалять. Для обнаружения выбросов используются простейшие способы, основанные на межквартильном расстоянии: выбросами считаются значения, не попавшие в диапазон $\left[\left(x_{25} - 1{,}5 \cdot (x_{75} - x_{25})\right), \left(x_{25} + 1{,}5 \cdot (x_{75} - x_{25})\right)\right]$. [13]

Анализ выбросов производят с помощью диаграмм размаха. Также их называют «ящики с усами». Они изображают одномерное распределение вероятностей.

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы. Несколько таких ящиков можно нарисовать бок о бок, чтобы визуально сравнивать одно распределение с другим; их можно располагать как горизонтально, так и вертикально. Расстояния между различными частями ящика позволяют определить степень разброса (дисперсии) и асимметрии данных и выявить выбросы.

2.2.2 Восстановление пропущенных значений

К сожалению, большинство статистических методов предполагает, что в ходе наблюдений были получены полностью укомплектованные матрицы, векторы и другие информационные структуры эксперимента. Поскольку на практике пропуски в данных все же являются повсеместным явлением, прежде чем начать аналитические изыскания, необходимо привести обрабатываемые таблицы к "каноническому" виду, т.е. либо удалить фрагменты объектов с недостающими элементами, либо заменить имеющиеся пропуски на некоторые разумные значения.

Несмотря на то, что книги по статистике скупо разбирают проблему исследования недостающих данных, в этой области существует впечатляющее множество подходов, методологий и их критических анализов. На практике процедура "борьбы с пропусками" обычно включает следующие шаги:

1. Идентификация недостающих данных.
2. Исследование закономерностей появления отсутствующих значений.
3. Формирование наборов данных, не содержащих пропуски (в результате удаления или замены соответствующих фрагментов).[14]

Необходимо признать, что идентификация недостающих данных является здесь единственным однозначным шагом. Анализ, почему данные отсутствуют, зависит от вашего понимания процессов, которые

воспроизводят экспериментальную информацию. Решение о способе устранения пропущенных значений также будет зависеть от вашей оценки того, какие процедуры приведут к самым надежным и точным результатам.

Исследование закономерностей появления отсутствующих значений производится для того чтобы получить представление о возможных механизмах возникновения пропущенных данных и о влиянии пропущенных данных на качество ответов на интересующие нас вопросы.

В частности, необходимо знать, какая доля данных пропущена, сосредоточены ли пропущенные данные в нескольких переменных или они широко распределены по всему набору данных, можно ли их считать случайными и позволяет ли ковариация пропущенных данных друг с другом или с наблюдаемыми данными обнаружить возможный механизм, лежащий в основе пропущенных значений.

Ответы на эти вопросы позволят определить, какие статистические методы лучше всего подходят для анализа данных.

Рациональный подход. При так называемом рациональном подходе (rational approach) при попытках замены или восстановления пропущенных значений используются математические или логические связи между переменными.

Для применения рационального подхода обычно требуется творческое мышление, наряду с должными навыками управления данными. Восстановление данных может быть точным или приблизительным.[14]

Анализ полных строк (построчное удаление). При анализе полных строк (complete-case analysis) работают только со строками без пропущенных значений. Многие широко используемые статистические пакеты по умолчанию используют построчное удаление (listwise/case-wise deletion) при работе с пропущенными данными. Такой подход настолько распространен, что многие аналитики, проводя регрессионный или дисперсионный анализы, могут даже не осознавать, что существует «проблема пропущенных данных», с которой нужно как-то справляться.

При построчном удалении данных предполагается, что пропуски полностью случайны (то есть полные строки – это случайная выборка из всего набора данных). Удаление всех строк с пропущенными данными может снизить статистическую мощность, уменьшив объем выборки.[14]

Метод множественного восстановления пропущенных данных (multiple imputation, MI) – это способ заполнения пропусков при помощи повторного моделирования. Множественное восстановление часто применяется для работы с пропущенными данными в сложных ситуациях. При этом подходе из существующего набора данных с пропущенными значениями создается несколько полных наборов данных (обычно от трех до десяти). Для замещения пропущенных значений в производных наборах данных используются методы Монте-Карло.

К каждому из производных наборов данных применяются стандартные статистические методы, а на основании их результатов формируются оценки окончательных результатов и доверительные интервалы, которые учитывают неопределенность, созданную пропущенными значениями.[14]

2.2.3 Факторный анализ

Для выявления взаимосвязей между значениями переменных используется факторный анализ. В его основе лежит предположение о том, что все переменные в имеющейся выборке зависят от меньшего числа латентных переменных и случайной ошибки.

В области факторного анализа существует два базисных понятия: фактор и нагрузка. Фактором называют неизвестную переменную, а нагрузкой – корреляцию между фактором и начальной переменной.

Возможности и задачи факторного анализа

Благодаря использованию факторного анализа могут быть решены две значимые для исследования проблемы: всестороннее и одновременно компактное описание объекта измерения. Также факторный анализ позволяет

выявлять латентные переменные-факторы, которые обязательно имеют линейные корреляции с исходными переменными.

К основным целям факторного анализа можно отнести следующие:

- определение корреляций между исследуемыми показателями;[16]
- сокращение признакового пространства исходной выборки.

В результате анализа происходит объединение в один фактор нескольких переменных, значение корреляции которых велико, вследствие этого создается достаточно простая для восприятия структура факторов из-за перераспределения дисперсии среди основных компонентов.

Вследствие объединения внутри каждого фактора уровень корреляции составляющих будет значительно выше, нежели уровень корреляции с объектами, принадлежащими другим факторам. Также в результате факторного анализа выделяются новые скрытые переменные, что позволяет сократить признаковое пространство исходной выборки, что может быть очень полезно при дальнейших исследованиях.

В качестве примера можно привести анализ оценок школьников. Допустим, в школе проводят занятие по таким предметам, как алгебра, биология, география, геометрия, иностранный язык, информатика, история, литература, обществознание, русский язык, физика и химия. Вероятнее всего, ученик, имеющий хорошие оценки по алгебре, имеет положительные оценки и по геометрии. То есть, если школьник имеет логический склад ума, его оценки по алгебре, геометрии и информатике, скорее всего, будут хорошими. Также ученики могут иметь способности к гуманитарным или естественным наукам. Если выборка будет содержать достаточное количество записей, то при проведении факторного анализа с выделением на три фактора, вероятнее всего, предметы разделятся на изучающие точные, гуманитарные и естественные науки.

Выделяют следующие виды факторного анализа:

- разведочный – применяется в случае, когда неизвестно о возможной структуре выделяемых факторов, их необходимом количестве и об их нагрузках;
- конфирматорный (подтверждающий) – осуществляется для подтверждения или опровержения предположения о количестве факторов и их нагрузках.[17]

Условия применения факторного анализа

Перед началом выполнения факторного анализа необходимо проверить обязательные условия:

- все исследуемые переменные должны быть описаны числовым типом данных;
- количество переменных должно как минимум в два раза меньше количества описанных наблюдений;
- исходные данные должны иметь однородную структуру без пропусков;
- распределение переменных в выборке должно быть симметричным;
- начальные данные должны иметь коррелирующие переменные для возможности осуществления факторного анализа.[16]

2.2.4 Кластерный анализ

Кластерный анализ применяется для исследования многомерных данных и их группировки в относительно однородные группы с использованием статистических методов.[18-19]

Спектр применений кластерного анализа очень широк: его используют в археологии, медицине, психологии, химии, биологии, государственном управлении, филологии, антропологии, маркетинге, социологии, геологии и других дисциплинах. Однако универсальность применения привела к появлению большого количества несовместимых терминов, методов и подходов, затрудняющих однозначное использование и непротиворечивую интерпретацию кластерного анализа.[19]

Кластерный анализ выполняет следующие основные задачи:

- Разработка типологии или классификации;
- Исследование полезных концептуальных схем группирования объектов;
- Порождение гипотез на основе исследования данных;
- Проверка гипотез или исследования для определения, действительно ли типы (группы), выделенные тем или иным способом, присутствуют в имеющихся данных.[19]

Независимо от предмета изучения применение кластерного анализа предполагает следующие этапы:

- Отбор выборки для кластеризации. Подразумевается, что имеет смысл кластеризовать только количественные данные;
- Определение множества переменных, по которым будут оцениваться объекты в выборке, то есть признакового пространства;
- Вычисление значений той или иной меры сходства (или различия) между объектами;
- Применение метода кластерного анализа для создания групп сходных объектов;
- Проверка достоверности результатов кластерного решения.[18]

Можно встретить описание двух фундаментальных требований предъявляемых к данным — однородность и полнота. Однородность требует, чтобы все кластеризуемые сущности были одной природы, описывались сходным набором характеристик. Если кластерному анализу предшествует факторный анализ, то выборка не нуждается в «ремонте» — изложенные требования выполняются автоматически самой процедурой факторного моделирования. В противном случае выборку нужно корректировать.

Цели кластеризации

- Понимание данных путём выявления кластерной структуры. Разбиение выборки на группы схожих объектов позволяет упростить дальнейшую

обработку данных и принятия решений, применяя к каждому кластеру свой метод анализа (стратегия «разделяй и властвуй»);

- Сжатие данных. Если исходная выборка избыточно большая, то можно сократить её, оставив по одному наиболее типичному представителю от каждого кластера;
- Обнаружение новизны (англ. novelty detection). Выделяются нетипичные объекты, которые не удаётся присоединить ни к одному из кластеров.[18]

В первом случае число кластеров стараются сделать поменьше. Во втором случае важнее обеспечить высокую степень сходства объектов внутри каждого кластера, а кластеров может быть сколько угодно. В третьем случае наибольший интерес представляют отдельные объекты, не вписывающиеся ни в один из кластеров.

Во всех этих случаях может применяться иерархическая кластеризация, когда крупные кластеры дробятся на более мелкие, те в свою очередь дробятся ещё мельче, и т. д. Такие задачи называются задачами таксономии. Результатом таксономии является древообразная иерархическая структура. При этом каждый объект характеризуется перечислением всех кластеров, которым он принадлежит, обычно от крупного к мелкому.

Методы кластеризации

Общепринятой классификации методов кластеризации не существует, но можно выделить ряд групп подходов (некоторые методы можно отнести сразу к нескольким группам и потому предлагается рассматривать данную типизацию как некоторое приближение к реальной классификации методов кластеризации):

- Вероятностный подход предполагает, что каждый рассматриваемый объект относится к одному из k классов. Некоторые авторы (например, А. И. Орлов) считают, что данная группа вовсе не относится к кластеризации и противопоставляют её под названием

«дискриминация», то есть выбор отнесения объектов к одной из известных групп (обучающих выборок);

- Подходы на основе систем искусственного интеллекта: весьма условная группа, так как методов очень много и методически они весьма различны;
- Логический подход – построение дендрограммы осуществляется с помощью дерева решений;
- Теоретико-графовый подход;
- Иерархический подход предполагает наличие вложенных групп (кластеров различного порядка). Алгоритмы в свою очередь подразделяются на агломеративные (объединительные) и дивизивные (разделяющие). По количеству признаков иногда выделяют монотетические и политетические методы классификации;
- Другие методы, не вошедшие в предыдущие группы.[20]
-

2.3 Описание инструмента решения

2.3.1 Скриптовый язык R

R — язык программирования для статистической обработки данных и работы с графикой. Изначально R был разработан сотрудниками статистического факультета Оклендского университета Россом Айхэкой и Робертом Джентлменом; язык и среда поддерживаются и развиваются организацией R Foundation.[21]

R широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистических программ.

В R используется интерфейс командной строки, хотя доступны и несколько графических интерфейсов пользователя, например пакет R Commander, RKWard, RStudio, Weka, Rapid Miner, KNIME, а также средства интеграции в офисные пакеты.

2.3.2 Особенности языка R

R поддерживает широкий спектр статистических и численных методов и обладает хорошей расширяемостью с помощью пакетов. Пакеты представляют собой библиотеки для работы специфических функций или специальных областей применения. В базовую поставку R включен основной набор пакетов, а всего по состоянию на 2017 год доступно более 11778 пакетов.[22]

Ещё одна особенность R – возможность создания качественной графики, которая может включать математические символы.

Достоинства среды R:

- бесплатная и кроссплатформенная;
- богатый арсенал статистических методов;
- качественная векторная графика;
- более 11000 проверенных пакетов;
- гибкая в использовании:
    - позволяет создавать и редактировать скрипты и пакеты,
    - взаимодействует с другими языками, такими: C, Java и Python,
    - может работать с форматами данных для SAS, SPSS и STATA;
- активное сообщество пользователей и разработчиков;
- регулярные обновления, хорошая документация и тех. поддержка.

Основным недостатком является небольшой объем информации на русском языке.

2.3.3 Форматы данных

С точки зрения статистики, данные принято делить на типы в зависимости от того, насколько близко их можно представить при помощи известной метафоры числовой прямой. Например, возраст человека легко представить таким образом, за тем исключением, что он не может быть отрицательным. Размер ботинок представить так уже сложнее, поскольку

между двумя соседним размерами, как правило, не бывает промежуточного значения. В то время как между двумя любыми числами на числовой прямой всегда можно найти нечто промежуточное. Зато размеры можно хотя бы расположить по возрастающей или по убывающей. А вот пол человека так представить уже совсем не получится: есть только два значения, и «промежуточного» просто не бывает. Мы, конечно, можем обозначить женский пол единицей, а мужской — нулём (или двойкой), но никакой числовой информации эти обозначения нести не будут — их даже нельзя отсортировать. Есть ещё и другие специальные виды данных, например, углы, географические координаты, даты и т. п., но все они так или иначе могут быть представлены с помощью чисел.

Таким образом, наиболее принципиальное различие между типами данных — это можно или нельзя их представить при помощи «обычных» чисел. Если нельзя, то такие данные принято называть категориальными. Статистические законы, а, значит, и статистические программы, работают с такими данными, только если заранее указан их тип. Остальные типы данных в разных книгах называют по-разному: числовые, счётные, порядковые или некатегориальные. Итак, основными типами данных являются:

- Числовые векторы;
- Факторы;
- Пропущенные данные;
- Матрицы;
- Списки.[22]

Глава 3. Разработка технологии оценки качества жизни

3.1 Алгоритм обработки данных

В ходе работы был разработан алгоритм (рис.1), согласно которому осуществлялся анализ исходных данных.

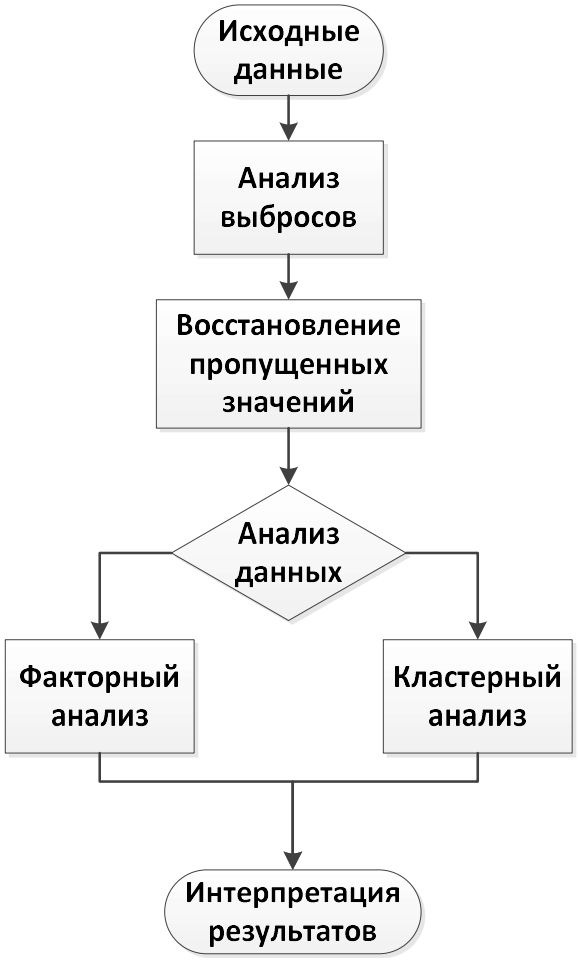


Рис. 1. Алгоритм анализа данных

3.2 Анализ выбросов

Для наглядности анализ выбросов производился для каждой методики отдельно. Для построения диаграмм размаха в R используется функция boxplot(). Скрипт для теста SF-36 Health Status Survey:

```
> boxplot(MySet$ОБСостЗдор_1,MySet$'ОБ Сост Здоровья 2',MySet$'Физич функц_1',MySet$'Физич функц_2',MySet$'ВлиянФиз Зд на функц_1',MySet$'ВлиянФиз Зд на функц_2',MySet$'ВлиянЭМ на функц_1',MySet$'ВлиянЭМ на функц_2',MySet$'Социальное функц_1',MySet$'Социальное функц_2',MySet$'Интенсивность боли_1',MySet$'Интенсивность боли_2',MySet$'Жизнестойкость_1',MySet$'Жизнестойкость_2',MySet$'Самооц психол. Зд_1',MySet$'Самооц психол. Зд_2')
```

Результат работы скрипта представлен на рис. 2.

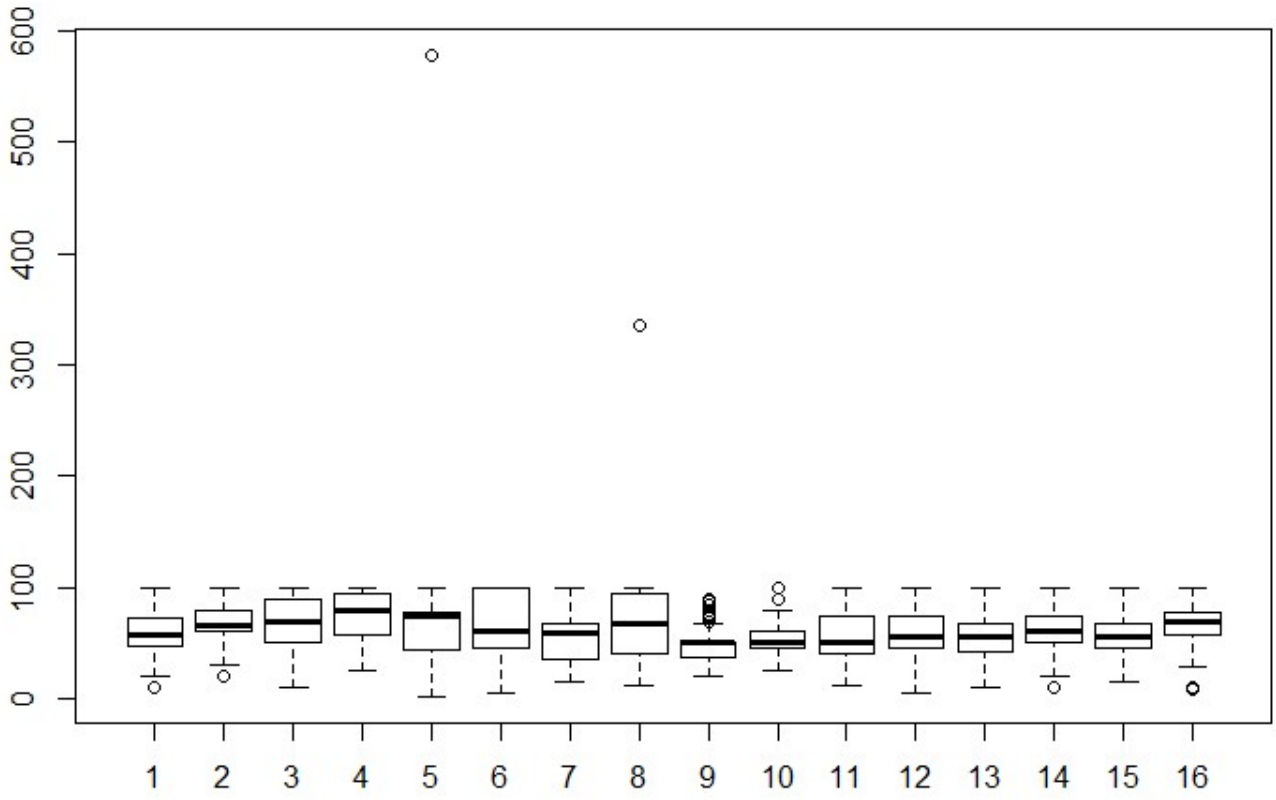


Рис. 2. Диаграммы размаха для теста SF-36

Показатели для этого теста варьируются от 0 до 100, поэтому точки со значениями больше 100 однозначно являются выбросами, вероятнее всего вызванные ошибками ввода данных. После удаления этих выбросов диаграммы приобретают следующий вид (рис.3):

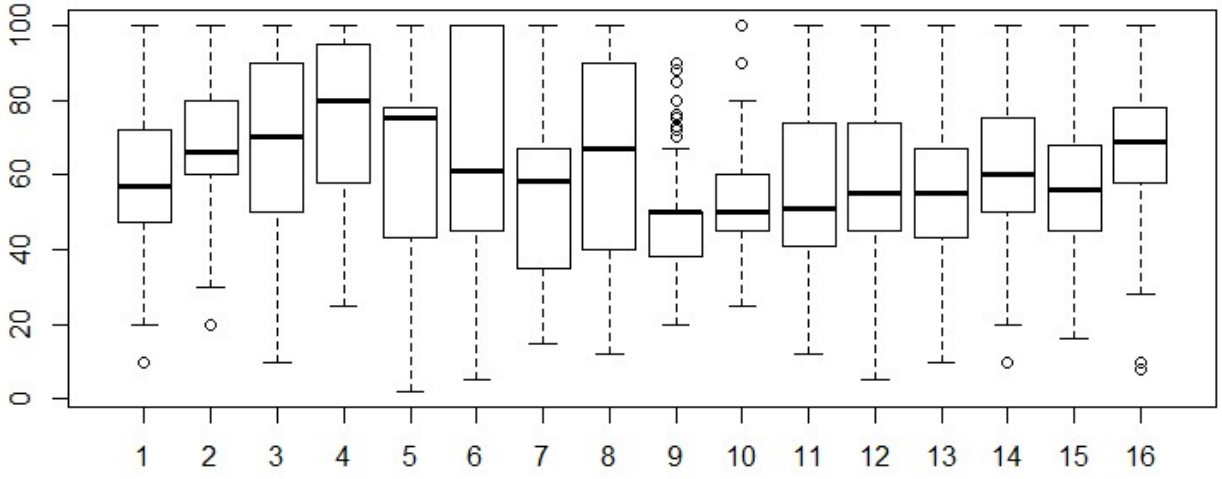


Рис. 3. Диаграммы размаха для теста SF-36 после удаления выбросов

В этой диаграмме также присутствуют потенциальные выбросы, но их так просто удалить мы не можем, так как значения показателей не противоречат методике тестирования.

Для теста цветовых восприятий Люшера также был написан скрипт:

```
> boxplot(MySet$'Л_Р 1',MySet$'Л_Р 2',MySet$'Л_С 1',MySet$'Л_С 2
',MySet$'Л_СО 1',MySet$'Л_СО 2',MySet$'ЛТ1',MySet$'ЛТ2',MySet$'С
т1',MySet$'Ст2',MySet$'М_Рнс 1',MySet$'М_Рнс 2')
```

Результат работы скрипта представлен на рис. 4.

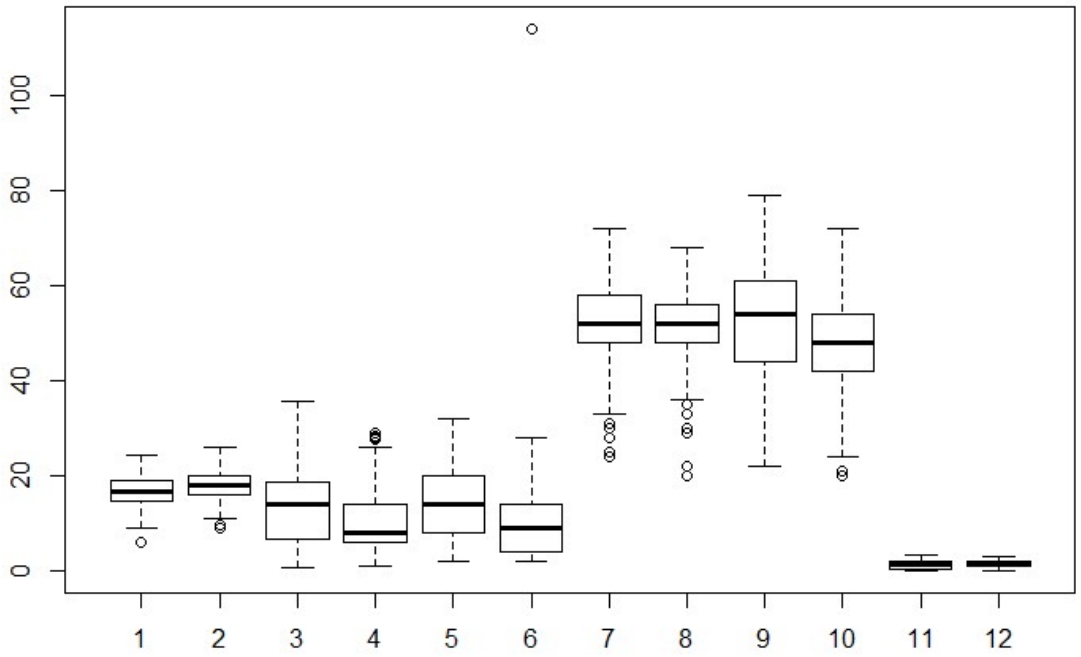


Рис. 4. Диаграммы размаха для теста Люшера

В данном случае верхняя точка тоже может быть удалена, так как имеет значение больше 100. После удаления этой точки остаются еще похожие на выбросы, но природа их происхождения не известна, поэтому эти «аномальные» точки не могут быть исключены (рис. 5).

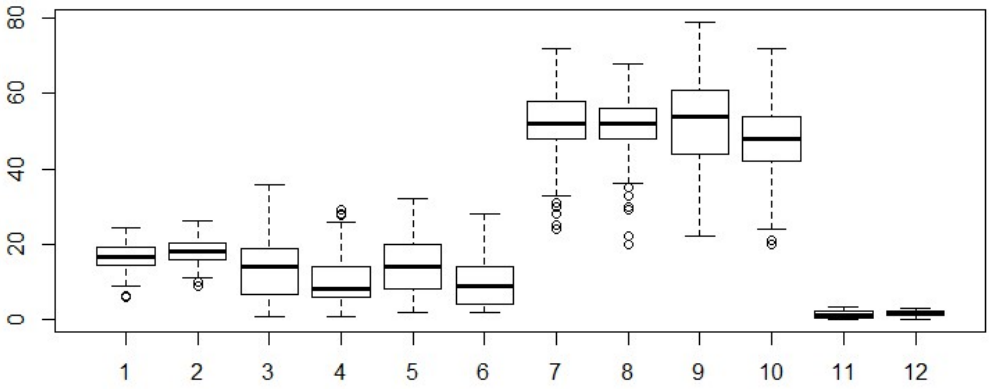


Рис. 5. Диаграммы размаха для теста Люшера после удаления выбросов

Для шкалы базисных убеждений Янов-Бульман значения показателей не могут превышать 6. Скрипт для данного теста выглядит следующим образом:

```
> boxplot(MySet1$'Благоскл Мира БУ',MySet1$'Доброта людей БУ',My
Set1$'Справедл мира БУ',MySet1$'Контролир Миара БУ',MySet1$'Вера
в случайность БУ',MySet1$'Ценность Я БУ',MySet1$'Степень самокон
троля БУ',MySet1$'Степень удачи БУ')
```

Полученные диаграммы изображены на рис. 6.

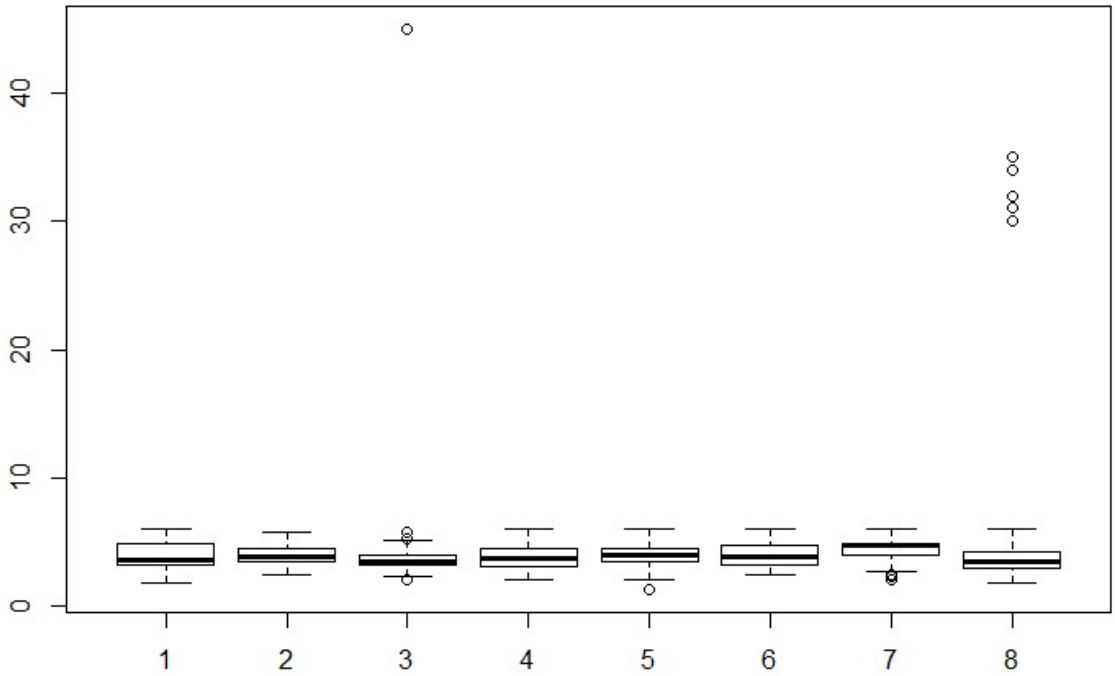


Рис. 6. Диаграммы размаха для шкалы базисных убеждений Янов-Бульман

Очевидно, что выбросы присутствуют в третьем и в восьмом показателях. После их удаления диаграммы приобретают следующий вид (рис. 7):

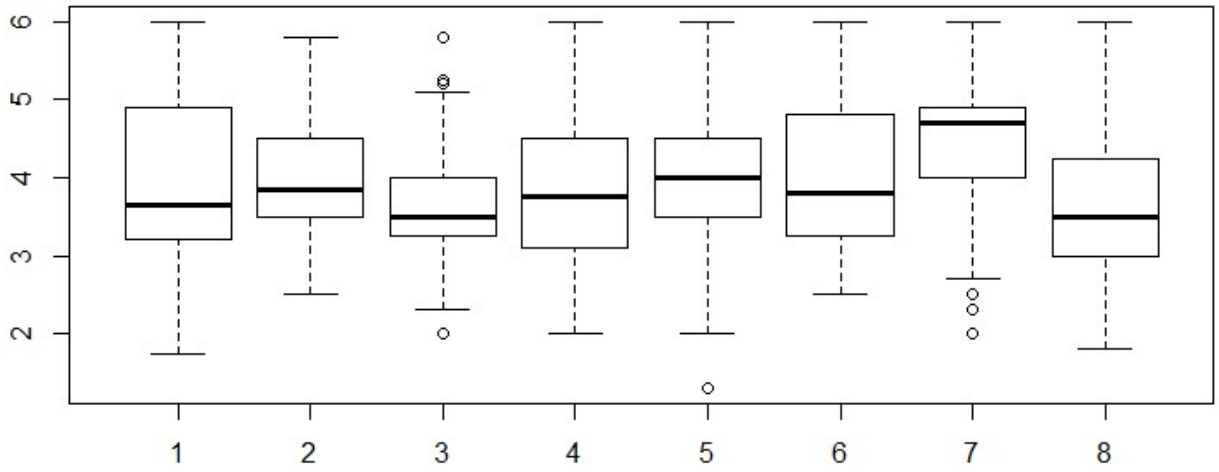


Рис. 7. Диаграммы размаха для шкалы базисных убеждений Янов-Бульман без выбросов

Скрипт для теста жизнестойкости Мадди:

```
> boxplot(MySet1$'Вовлеченность Жизнестойкость',MySet1$'Контроль Жизнестойкость',MySet1$'Риск Жизнестойкость',MySet1$'Общая Жизнестойкость',MySet1$'ВыражЖИЗН')
```

В результате выполнения скрипта аномальных выбросов не обнаружено (рис.8).

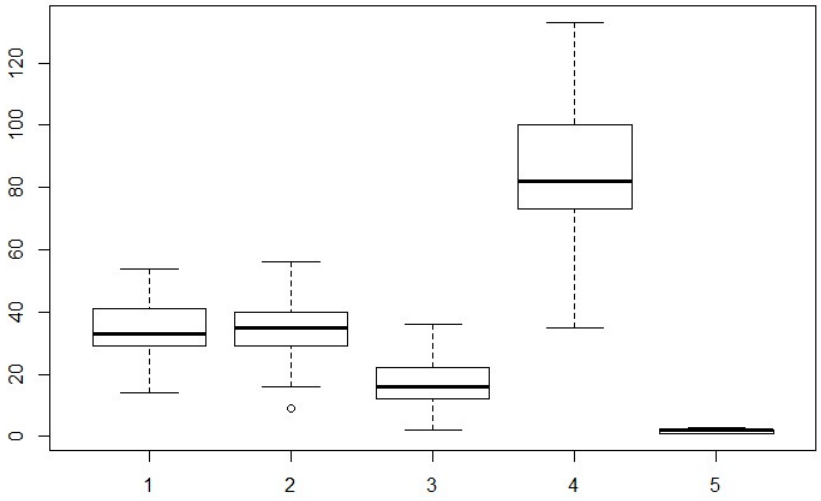


Рис. 8. Диаграммы размаха для теста жизнестойкости Мадди

В результате анализа выбросов для опросника самоорганизации деятельности аномальных значений также обнаружено не было (рис. 9). Скрипт для этой методики:

```
> boxplot(MySet1$'Планир.',MySet1$'Целеустр.',MySet1$'Наст.',MySet1$'Фиксац.',MySet1$'Самоорг.',MySet1$'Будущ.',MySet1$'ОСД Сумма',MySet1$'ОСД_ИНД',MySet1$'ИНД_ЦЕЛЕУСТР',MySet1$'ИНД_РАЦИОН.')
```

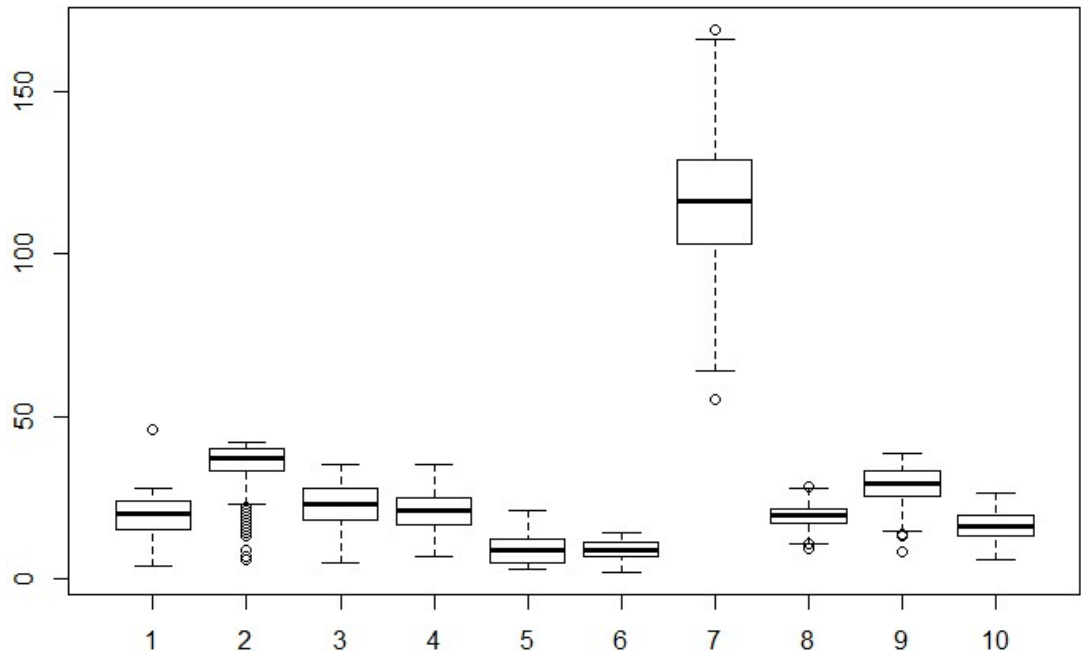


Рис. 9. Диаграммы размаха для опросника самоорганизации деятельности

3.3 Визуализация и восстановление пропущенных значений

Скриптовый язык R позволяет выделить полностью укомплектованные строки в массиве данных. Для этого используется следующий скрипт:

```
> MySet1[complete.cases(MySet1),]
```

Результат работы данного скрипта представлен на рис. 10.

```
> MySet1[complete.cases(MySet1),]
# A tibble: 185 x 53
   `№`     `Д-З` ОБСостЗдор_1 `ОБ Сост Здоровь~ `Физич функц_1` `Физич функц_2` `ВлиянФиз Зд на ф~ `ВлиянФиз Зд на ф~
   <chr>   <dbl>        <dbl>             <dbl>           <dbl>           <dbl>             <dbl>             <dbl>
 1 Person~     7.           55.               62.             50.             55.               50.               56.
 2 Person~     7.           50.               64.             55.             67.               75.               85.
 3 Person~     7.           40.               55.             45.             54.               25.               45.
 4 Person~     7.           55.               55.             64.             68.               35.               66.
 5 Person~     7.           50.               64.             55.             67.               75.               85.
 6 Person~     7.           40.               55.             45.             54.               25.               45.
 7 Person~     7.           55.               55.             64.             68.               35.               66.
 8 Person~     7.           55.               62.             50.             55.               50.               56.
 9 Person~     7.           50.               64.             55.             67.               75.               85.
10 Person~     7.           45.               80.             90.             75.               75.               95.
# ... with 175 more rows, and 45 more variables: `ВлиянЭМ на функц_1` <dbl>, `ВлиянЭМ на функц_2` <dbl>, `Социальное
#   функц_1` <dbl>, `Социальное функц_2` <dbl>, `Интенсивность боли_1` <dbl>, `Интенсивность боли_2` <dbl>,
#   Жизнестойкость_1 <dbl>, Жизнестойкость_2 <dbl>, `Самооц психол. Зд_1` <dbl>, `Самооц психол. Зд_2` <dbl>, `Л_Р
#   1` <dbl>, `Л_Р 2` <dbl>, `Л_С 1` <dbl>, `Л_С 2` <dbl>, `Л_СО 1` <dbl>, `Л_СО 2` <dbl>, ЛТ1 <dbl>, ЛТ2 <dbl>,
#   Ст1 <dbl>, Ст2 <dbl>, `М_Рнс 1` <dbl>, `М_Рнс 2` <dbl>, `Благоскл Мира БУ` <dbl>, `Доброта людей БУ` <dbl>,
#   `Справедл мира БУ` <dbl>, `Контролир Миара БУ` <dbl>, `Вера в случайность БУ` <dbl>, `Ценность Я БУ` <dbl>,
#   `Степень самоконтроля БУ` <dbl>, `Степень удачи БУ` <dbl>, `Вовлеченность Жизнестойкость` <dbl>, `Контроль
#   Жизнестойкость` <dbl>, `Риск Жизнестойкость` <dbl>, `Общая Жизнестойкость` <dbl>, ВыражЖизН <dbl>, Планир. <dbl>,
#   Целеустр. <dbl>, Наст. <dbl>, Фиксац. <dbl>, Самоорг. <dbl>, Будущ. <dbl>, `ОСД Сумма` <dbl>, ОСД_ИНД <dbl>,
#   ИНД_ЦЕЛЕУСТР <dbl>, ИНД_РАЦИОН. <dbl>
```

Рис. 10. Укомплектованные строки массива данных

Из 663 строк полностью заполненными оказались лишь 185. Для проверки результатов, выполним скрипт, который показывает строки, в которых имеется пропуск хотя бы одного значения:

```
> MySet1[!complete.cases(MySet1),]
```

```
> MySet1[!complete.cases(MySet1),]
# A tibble: 478 x 53
   `№`     `Д-З` ОБСостЗдор_1 `ОБ Сост Здоровь~ `Физич функц_1` `Физич функц_2` `ВлиянФиз Зд на ф~ `ВлиянФиз Зд на ф~
   <chr>   <dbl>        <dbl>             <dbl>           <dbl>           <dbl>             <dbl>             <dbl>
 1 Person~     7.           45.               55.             90.             90.               75.               88.
 2 Person~     7.           55.               69.             80.             90.               75.               80.
 3 Person~     7.           55.               67.             40.             58.               25.               49.
 4 Person~     7.           55.               59.             90.             94.               75.               80.
 5 Person~     7.           50.               66.             70.             89.               NA                56.
 6 Person~     7.           45.               58.             44.             59.               NA                45.
 7 Person~     7.           92.              100.             70.             98.               NA                78.
 8 Person~     7.           55.               67.             40.             58.               25.               49.
 9 Person~     7.           55.               59.             90.             94.               75.               80.
10 Person~     7.           50.               66.             70.             89.               NA                56.
# ... with 468 more rows, and 45 more variables: `ВлиянЭМ на функц_1` <dbl>, `ВлиянЭМ на функц_2` <dbl>, `Социальное
#   функц_1` <dbl>, `Социальное функц_2` <dbl>, `Интенсивность боли_1` <dbl>, `Интенсивность боли_2` <dbl>,
#   Жизнестойкость_1 <dbl>, Жизнестойкость_2 <dbl>, `Самооц психол. Зд_1` <dbl>, `Самооц психол. Зд_2` <dbl>, `Л_Р
#   1` <dbl>, `Л_Р 2` <dbl>, `Л_С 1` <dbl>, `Л_С 2` <dbl>, `Л_СО 1` <dbl>, `Л_СО 2` <dbl>, ЛТ1 <dbl>, ЛТ2 <dbl>,
#   Ст1 <dbl>, Ст2 <dbl>, `М_Рнс 1` <dbl>, `М_Рнс 2` <dbl>, `Благоскл Мира БУ` <dbl>, `Доброта людей БУ` <dbl>,
#   `Справедл мира БУ` <dbl>, `Контролир Миара БУ` <dbl>, `Вера в случайность БУ` <dbl>, `Ценность Я БУ` <dbl>,
#   `Степень самоконтроля БУ` <dbl>, `Степень удачи БУ` <dbl>, `Вовлеченность Жизнестойкость` <dbl>, `Контроль
#   Жизнестойкость` <dbl>, `Риск Жизнестойкость` <dbl>, `Общая Жизнестойкость` <dbl>, ВыражЖизН <dbl>, Планир. <dbl>,
#   Целеустр. <dbl>, Наст. <dbl>, Фиксац. <dbl>, Самоорг. <dbl>, Будущ. <dbl>, `ОСД Сумма` <dbl>, ОСД_ИНД <dbl>,
#   ИНД_ЦЕЛЕУСТР <dbl>, ИНД_РАЦИОН. <dbl>
```

Рис. 11. Строки с пропусками

Таким образом, исходные данные содержат 478 строк с пропусками, которые необходимо восстановить.

Сначала выполним визуализацию пропущенных значений, чтобы наглядно посмотреть, в каких тестах имеются пропуски. Для этого выполним следующий скрипт:

```
> aggr(X1_sf_36,prop=TRUE,numbers=TRUE)
```

На рис. 12 представлены пропущенные значения для теста SF-36 Health Status Survey.

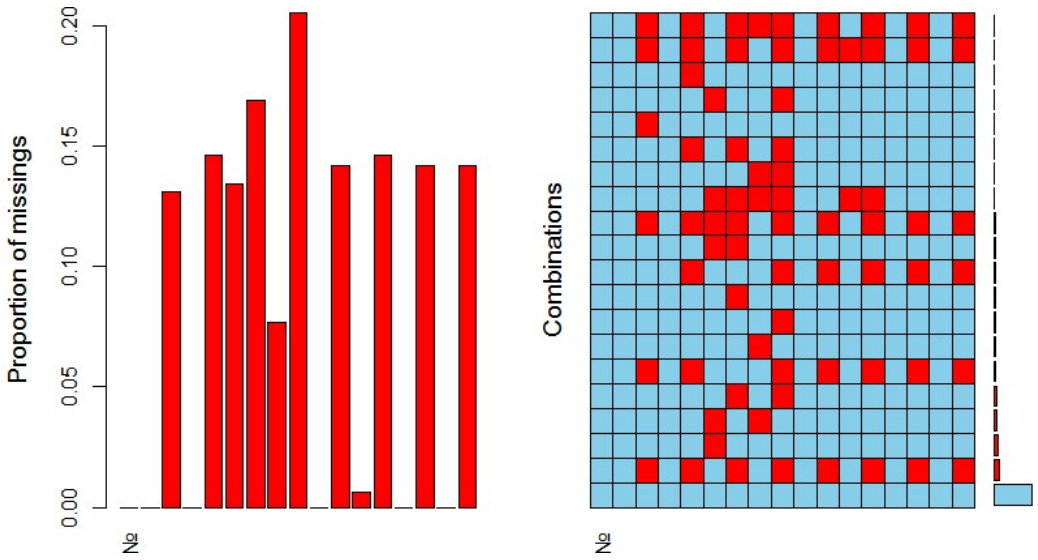


Рис. 12. Визуализация пропущенных значений для теста SF-36

На диаграмме слева представлено количество пропущенных значений для каждого показателя в отдельности, а справа – для комбинации показателей.

Для цветового теста Люшера диаграмма выглядит следующим образом (рис. 13):

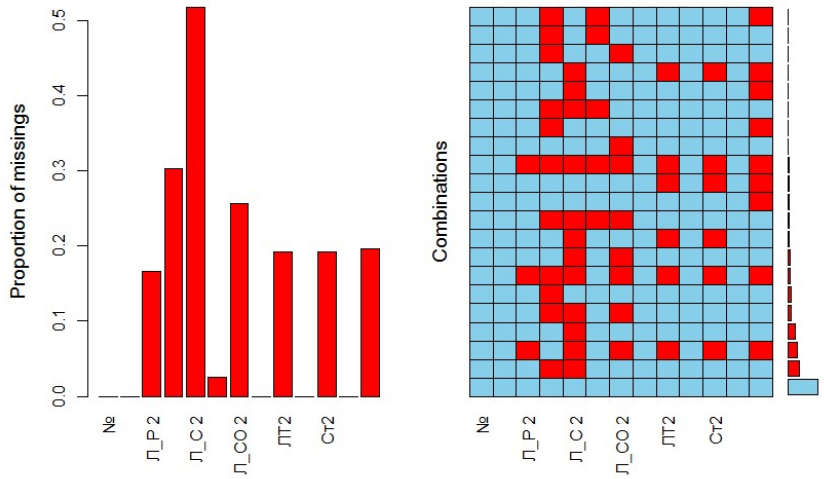


Рис. 13. Визуализация пропущенных значений для теста Люшера

Для теста жизнестойкости Мадди в адаптации Леонтьева и для шкалы базисных убеждений Янов-Бульман строк с пропущенными значениями выявлено не было.

Опросник самоорганизации деятельности содержит небольшое количество пропусков (рис. 14):

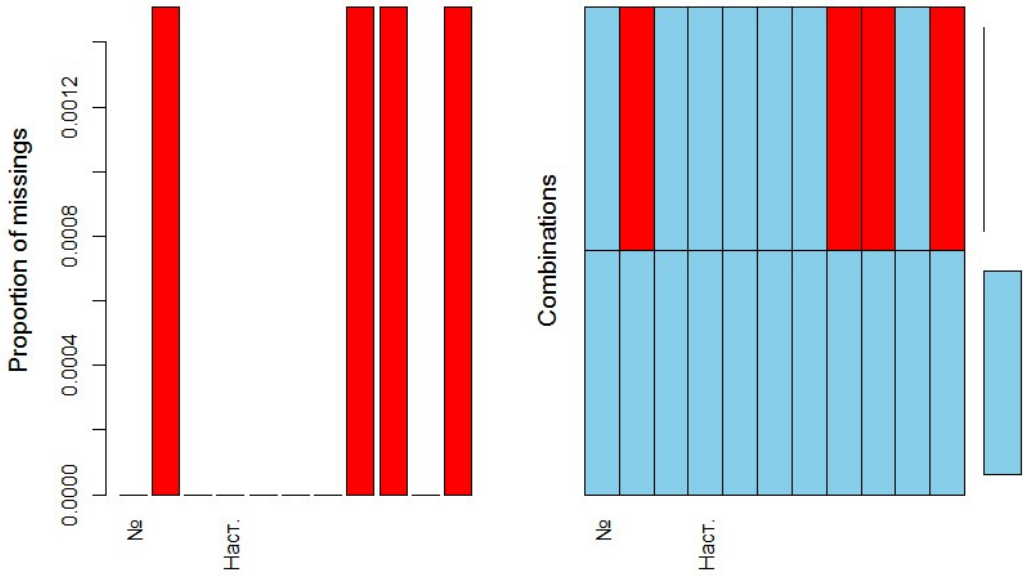


Рис. 14. Визуализация пропущенных значений для теста ОСД

Теперь перейдем к восстановлению пропущенных значений в тестах с пропусками. Для этого необходимо выполнить следующую последовательность скриптов:

```
> imp <- mice(MySet1)
> MySet2 <- complete(imp, action=1)
```

На рис. 15 приведен фрагмент выборки до выполнения скрипта:

	№	ОБСостЗдор_1	ОБ Сост Здоровья 2	Физич функц_1	Физич функц_2	ВлиянФиз Зд на функц_1	ВлиянФиз Зд на функц_2	ВлиянЭМ на функц_1	ВлиянЭМ на функц_2	Социальное функц_1	Социальное функц_2	Ин бо
9	Person 9	50	66	70	89	NA	56	NA	35	50	67	
10	Person 10	45	58	44	59	NA	45	38	50	40	54	
11	Person 11	92	100	70	98	NA	78	34	67	88	100	
12	Person 12	50	64	55	67	75	85	100	100	50	65	
13	Person 13	40	55	45	54	25	45	67	70	50	60	
14	Person 14	55	67	40	58	25	49	34	54	25	30	
15	Person 15	55	59	90	94	75	80	67	72	25	34	
16	Person 16	55	55	64	68	35	66	35	45	50	55	
17	Person 17	50	66	70	89	NA	56	NA	35	50	67	
18	Person 18	45	58	44	59	NA	45	38	50	40	54	
19	Person 19	45	55	90	90	75	88	67	78	25	49	
20	Person 20	55	69	80	90	75	80	67	71	38	55	
21	Person 21	55	62	50	55	50	56	67	70	50	70	
22	Person 22	50	64	55	67	75	85	100	100	50	65	

Рис. 15. Фрагмент выборки с пропусками

Пропущенные значения отражаются символами NA (not available). После выполнения скрипта выборка приобрела следующий вид (рис.16):

	№	ОБСостЗдор_1	ОБ Сост Здоровья 2	Физич функц_1	Физич функц_2	ВлиянФиз Зд на функц_1	ВлиянФиз Зд на функц_2	ВлиянЭМ на функц_1	ВлиянЭМ на функц_2	Социальное функц_1	Социальное функц_2	Ин бо
9	Person 9	50	66	70	89	78	56	67	35	50	67	
10	Person 10	45	58	44	59	35	45	38	50	40	54	
11	Person 11	92	100	70	98	50	78	34	67	88	100	
12	Person 12	50	64	55	67	75	85	100	100	50	65	
13	Person 13	40	55	45	54	25	45	67	70	50	60	
14	Person 14	55	67	40	58	25	49	34	54	25	30	
15	Person 15	55	59	90	94	75	80	67	72	25	34	
16	Person 16	55	55	64	68	35	66	35	45	50	55	
17	Person 17	50	66	70	89	76	56	30	35	50	67	
18	Person 18	45	58	44	59	78	45	38	50	40	54	
19	Person 19	45	55	90	90	75	88	67	78	25	49	
20	Person 20	55	69	80	90	75	80	67	71	38	55	
21	Person 21	55	62	50	55	50	56	67	70	50	70	

Рис. 16. Фрагмент выборки с восстановленными значениями

Восстановленные значения соответствуют интервалам шкал для всех методик.

3.4 Факторный анализ

В среде программирования RStudio факторный анализ осуществляется с использованием функции *factanal()*. Выделим для исходных данных 5 факторов:

```
> factanal(MySet2[,3:53],5)
```

Данная функция на выходе строит таблицу «уникальностей», то есть долей общей дисперсии, вносимой каждым показателем (рис. 17).

```
Uniquenesses:
     ОБСостЗдор_1  ОБ Сост Здоровья 2       Физич функц_1       Физич функц_2 
             0.781               0.854               0.803               0.801 
ВлиянФиз Зд на функц_1 ВлиянФиз Зд на функц_2       ВлиянЭМ на функц_1 ВлиянЭМ на функц_2 
             0.848               0.833               0.812               0.770 
Социальное функц_1  Социальное функц_2 Интенсивность боли_1 Интенсивность боли_2 
             0.952               0.972               0.808               0.792 
  Жизнестойкость_1    Жизнестойкость_2  Самооц психол. Зд_1  Самооц психол. Зд_2 
             0.612               0.601               0.740               0.744 
           Л_Р 1               Л_Р 2               Л_С 1               Л_С 2 
             0.584               0.820               0.525               0.844 
          Л_СО 1              Л_СО 2                 ЛТ1                 ЛТ2 
             0.521               0.782               0.669               0.717 
             Ст1                 Ст2               М_Рнс 1             М_Рнс 2 
             0.797               0.853               0.597               0.771 
 Благоскл Мира БУ    Доброта людей БУ  Справедл мира БУ  Контролир Миара БУ 
             0.167               0.376               0.227               0.191 
 Вера в случайность БУ  Ценность Я БУ Степень самоконтроля БУ Степень удачи БУ 
             0.567               0.201               0.639               0.334 
Вовлеченность Жизнестойкость Контроль Жизнестойкость Риск Жизнестойкость Общая Жизнестойкость 
             0.005               0.005               0.005               0.005 
          ВыражЖизН             Планир.              Целеустр.               Наст. 
             0.147               0.996               0.989               0.987 
            Фиксац.            Самоорг.                                        
             0.987               0.979                                        
```

Рис. 17. Таблица уникальностей

Затем идет таблица нагрузок, которая показывает значение корреляции отдельных параметров с выделенными факторами (рис. 18). Чтобы показатель оказывал влияние на фактор, необходимо, чтобы значение корреляции по модулю превышало 0,5.

Из этой таблицы видно, что первый фактор объединяет параметры шкалы базисных убеждений Янов-Бульман (корреляция со всеми показателями более 0,5). Второй фактор имеет взаимосвязь с показателями теста жизнестойкости Мадди. Третий – основывается на некоторых показателях теста SF-36 Health Status Survey, четвертый основывается на шкале Контроль теста жизнестойкости, а пятый имеет максимальное по модулю значение корреляции со шкалой Вовлеченность того же теста.

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5
ОбСостЗдор_1	0.163	0.248	0.329		0.125
ОБ Сост Здоровья 2	0.223		0.267		0.115
Физич функц_1		0.178	0.400		
Физич функц_2			0.438		
ВлиянФиз Зд на функц_1	-0.113		0.352		
ВлиянФиз Зд на функц_2			0.395		
ВлиянЭМ на функц_1			0.427		
ВлиянЭМ на функц_2			0.451	0.161	
Социальное функц_1			0.149	0.100	0.124
Социальное функц_2			0.163		
Интенсивность боли_1			0.423		
Интенсивность боли_2			0.443		
Жизнестойкость_1	0.176	0.126	0.580		
Жизнестойкость_2			0.623		
Самооц психол. Зд_1	0.112		0.489		
Самооц психол. Зд_2			0.494		
Л_Р 1	0.475	0.385	0.130	0.148	
Л_Р 2	0.259	0.200	0.156	0.218	
Л_С 1	-0.316	-0.399	-0.197		
Л_С 2	-0.335		-0.150	-0.130	
Л_СО 1	-0.491	-0.381	-0.201	-0.223	
Л_СО 2	-0.279	-0.134	-0.132	-0.323	
ЛТ1	-0.210	-0.182	-0.497		
ЛТ2	-0.146	-0.158	-0.480		
Ст1	-0.121	-0.109	-0.403		
Ст2			-0.353	-0.134	
М_Рнс 1	0.434	0.350	0.227	0.178	
М_Рнс 2	0.228	0.282	0.267	0.143	
Благоскл Мира БУ	0.899	0.145			
Доброта людей БУ	0.776	0.129			
Справедл мира БУ	0.868	0.101			
Контролир Миара БУ	0.895				
Вера в случайность БУ	-0.645		0.100		
Ценность Я БУ	0.875	0.166			
Степень самоконтроля БУ	0.573			0.128	
Степень удачи БУ	0.806	0.110			
Вовлеченность Жизнестойкость	0.394	0.777	0.130	0.131	-0.450
Контроль Жизнестойкость		0.451		0.857	0.237
Риск Жизнестойкость	0.281	0.882			0.370
Общая Жизнестойкость	0.279	0.863		0.414	
ВыражЖизн	0.311	0.807		0.316	
Планир.					
Целеустр.					
Наст.					
Фиксац.					
Самоорг.		-0.120			

Рис. 18. Таблица нагрузок

Следующая таблица – таблица долей общей дисперсии (рис. 19).

	Factor1	Factor2	Factor3	Factor4	Factor5
SS loadings	6.999	4.059	3.944	1.472	0.526
Proportion Var	0.202	0.188	0.116	0.088	0.086
Cumulative Var	0.202	0.390	0.506	0.594	0.680

Рис. 19. Таблица долей общей дисперсии

Таким образом, выделенные факторы описывают 68% разброса, при этом первый фактор объясняет 20,2%, второй – 18,8%, третий – 11,6%, четвертый – 8,8%, а пятый – 8,6%.

3.5 Кластерный анализ

3.5.1 Иерархический подход

Существуют различные виды кластерного анализа, но наиболее распространенными являются иерархические методы. Для построения дерева кластеризации используется следующий набор скриптов:

```
> a1<-MySet2[seq(1,nrow(MySet2),10),]
> a1.dist<-daisy(a1[,3:53])
> a1.h<-hclust(a1.dist, method="ward.D")
> plot(a1.h, method="both.sides"))
```

Для построения дерева использовалась каждая десятая строка, иначе это дерево было бы слишком велико для восприятия (рис. 20).

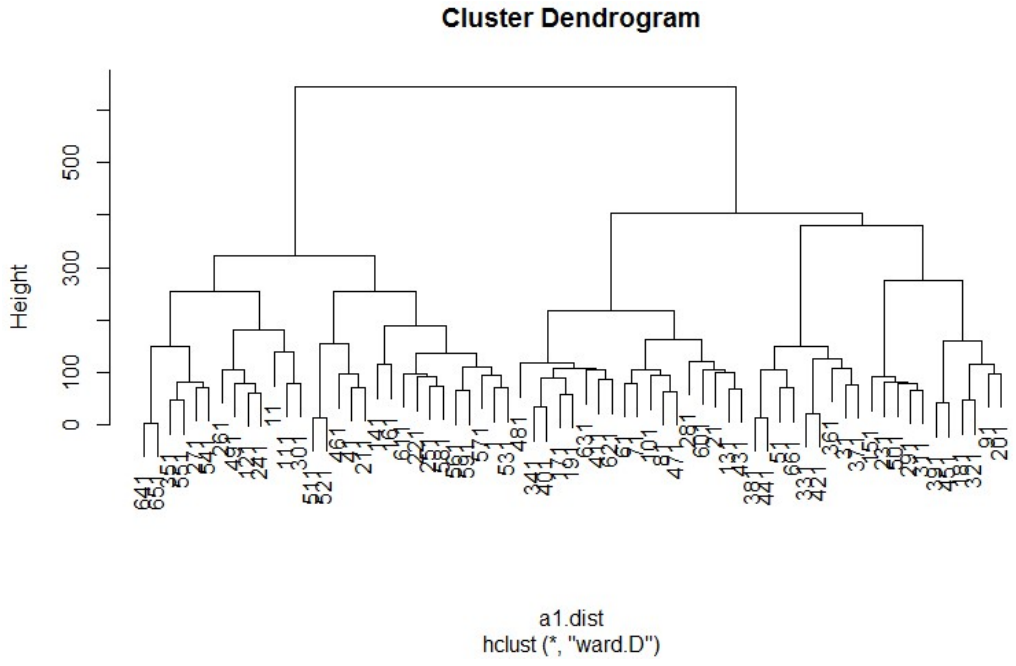


Рис. 20. Иерархическая кластеризация

По данному дереву можно выделить 4 кластера.

3.5.2 Метод k-средних

Найдем разбивку на 4 кластера по методу k-средних:

```
> c1<-kmeans(mydata,4)
> clusplot(mydata, c1$cluster, main='2D representation of the Cl
uster solution', color=TRUE, shade=TRUE, labels=2, lines=0)
```

В результате было получено следующее разбиение (рис. 21):

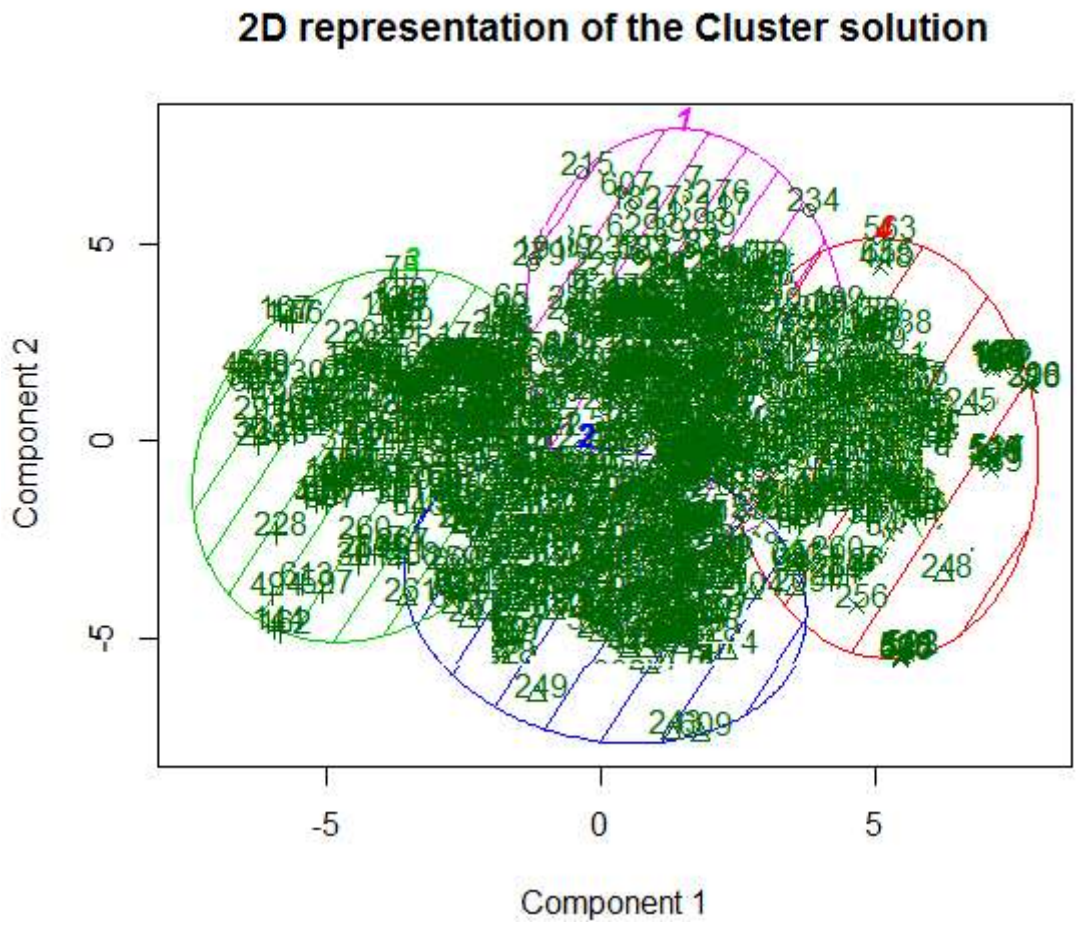


Рис. 21. Кластеризация методом k-средних

Кластерный анализ позволил выявить 4 достаточно обособленные группы людей, хотя и заметны пересечения.

3.5.3 Оценка качества кластеризации

Для оценки качества кластеризации был использован метод оценки силуэтов.

Пусть a – среднее расстояние от данного объекта до всех других объектов, лежащих внутри этого же кластера, а b – среднее расстояние от

данного объекта до всех других объектов, лежащих внутри ближайшего кластера, тогда силуэт объекта находится по следующей формуле:

$$s = \frac{b-a}{\min(a,b)}.$$

Индекс силуэта для кластера вычисляется как среднее арифметическое силуэтов всех объектов, входящих в кластер. В лучшем случае индекс силуэта должен быть равен 1.

Для нахождения индекса силуэта в R используется следующий скрипт:

```
> si <- silhouette(kmeans(mydata, 4))
```

Полученные значения индексов силуэта приведены в таблице 4.

Таблица 4. Значения индексов силуэтов

i	1 «розовый»	2 «синий»	3 «зеленый»	4 «красный»
S_i	0,563	0,731	0,811	0,602

3.6 Интерпретация результатов

В ходе анализа выбросов были построены диаграммы размаха, изучены «аномальные» точки в контексте исследуемых показателей и принято решение о возможности их удаления.

Была восстановлена однородность исходной выборки, используя метод множественного восстановления данных, который является достаточно сложным и ресурсоемким, но в то же время дает довольно точный результат. Суть метода заключается в генерации нескольких значений пропущенной величины и выборе наиболее подходящей, вместо того чтобы заменять одним значением.

В процессе выполнения факторного анализа было выявлено пять латентных переменных, которые позволили сократить признаковое пространство, состоящее из 51 переменной. Полученные пять факторов смогли описать 68% разброса. Они имели наибольшее значение корреляции с показателями из трех методик.

В результате проведения кластерного анализа были сформированы четыре группы людей со схожими результатами тестирования. Каждая группа была описана выявленными факторами.

Таким образом, все люди могут быть описаны данными трёх тестов оценки качества жизни: SF-36 Health Status Survey; шкала базисных убеждений Янов-Бульман; тест жизнестойкости Мадди.

В дальнейшем у опрашиваемых пациентов из списков методик можно исключить цветовой тест Люшера и опросник самоорганизации деятельности. Это позволит сократить время заполнения и, возможно, повысит качество заполнения, то есть люди не будут отвечать «хоть бы что» от усталости от заполнения большого объема тестов.

В результате проведения кластерного анализа было выделено четыре группы людей, которые совокупно описываются пятью факторами, выявленными в результате факторного анализа.

Для описания первой группы людей, изображенной в «розовом» кластере, используются второй и третий факторы, которые коррелируют с тестами жизнестойкости Мадди и SF-36 Health Status Survey.

Второй «синий» кластер может быть описан первым и четвертым факторами, то есть для определения принадлежности к этому кластеру необходимо знать результаты опроса по шкале базисных убеждений Янов-Бульман и значение шкалы Вовлеченность из теста жизнестойкости Мадди.

Третья группа людей («зеленый» кластер) описывается первым и третьим факторами, то есть основными методиками для описания этой группы являются шкала базисных убеждений Янов-Бульман и тест SF-36 Health Status Survey.

Четвертый выявленный «красный» кластер может быть описан третьим и пятым факторами, то есть методиками SF-36 Health Status Survey и отдельной шкалой Контроль теста жизнестойкости Мадди.

Оценка качества кластеризации показала, что полученный результат является корректным.

Глава 4. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения

4.1.1 Потенциальные потребители результатов исследования

Для анализа потребителей результатов исследования необходимо рассмотреть целевой рынок и провести его сегментирование.

Целевой рынок – сегменты рынка, на котором будет продаваться в будущем разработка. В свою очередь, сегмент рынка – это особым образом выделенная часть рынка, группы потребителей, обладающих определенными общими признаками.

Сегментирование – это разделение покупателей на однородные группы, для каждой из которых может потребоваться определенный товар (услуга). Можно применять географический, демографический, поведенческий и иные критерии сегментирования рынка потребителей, возможно применение их комбинаций с использованием таких характеристик, как возраст, пол, национальность, образование, любимые занятия, стиль жизни, социальная принадлежность, профессия, уровень дохода.

В зависимости от категории потребителей (коммерческие организации, физические лица) необходимо использовать соответствующие критерии сегментирования.

Таблица 5. Сегментирование рынка

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="3">Область применеия</td></tr>
<tr><td>Социология</td><td>Психология</td><td>Медицина</td></tr>
<tr><td rowspan="3">Использование технологии</td><td>В пределах страны</td><td colspan="3">Долгосрочная перспектива</td></tr>
<tr><td>В пределах региона</td><td>Долгосрочная перспектива</td><td colspan="2">Краткосрочная перспектива</td></tr>
<tr><td>В пределах города</td><td>Краткосрочная перспектива</td><td colspan="2">На этапе внедрения</td></tr>
</table>

4.1.2 Анализ конкурентных технических решений

Детальный анализ конкурирующих разработок, существующих на рынке, необходимо проводить систематически, поскольку рынки пребывают в постоянном движении. Такой анализ помогает вносить коррективы в научное исследование, чтобы успешнее противостоять своим соперникам.

Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения позволяет провести оценку сравнительной эффективности научной разработки и определить направления для ее будущего повышения.

Для этого была использована оценочная карта (таблица 2).

Таблица 6. Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	**Вес крите-рия**	**Баллы**			**Конкуренто-способность**		
		$Б_ф$	$Б_{к1}$	$Б_{к2}$	$К_ф$	$К_{к1}$	$К_{к2}$
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
1. Удобство в заполнении методик	0,1	5	4	5	0,5	0,4	0,5
2. Продолжительность обработки массива данных	0,2	5	3	4	1	0,6	0,8
3. Качество полученных результатов	0,1	4	5	4	0,4	0,5	0,4
4. Возможности технологии	0,05	5	4	5	0,25	0,2	0,25
5. Простота алгоритма	0,1	5	5	5	0,5	0,5	0,5
6. Пути внедрения	0,05	4	4	5	0,2	0,2	0,25
Экономические критерии оценки эффективности							
1. Конкурентоспособность технологии	0,1	4	5	3	0,4	0,5	0,3
2. Цена технологии	0,15	5	5	4	0,75	0,75	0,6
3. Срок внедрения	0,1	3	3	2	0,3	0,3	0,2
4. Наличие сертификации разработки	0,05	4	5	5	0,2	0,25	0,25
Итого	**1**				**4,5**	**4,2**	**4,05**

Для оценки ресурсоэффективности были выбраны следующие критерии: удобство в заполнении методик, продолжительность обработки массива данных, качество полученных результатов, возможности технологии, простота алгоритма, пути внедрения технологии. Наиболее

значимым критерием является продолжительность обработки массива данных.

Для оценки эффективности были выбраны следующие экономические критерии: конкурентоспособность технологии, цена, срок внедрения и наличие сертификации разработки.

Результаты анализа выявили, что разработанная технология выгодно отличается от конкурентов благодаря быстрой скорости обработки данных, удобству заполнения, большим возможностям при относительной простоте алгоритма.

4.1.3 Технология QuaD

Технология QuaD (QUality ADvisor) представляет собой гибкий инструмент измерения характеристик, описывающих качество новой разработки и ее перспективность на рынке и позволяющие принимать решение целесообразности вложения денежных средств в научно-исследовательский проект.

В основе технологии QuaD лежит нахождение средневзвешенной величины следующих групп показателей: показателей оценки коммерческого потенциала разработки и показателей оценки качества разработки.

Анализ проводится в виде оценочной карты (таблица 3).

Таблица 7. Оценочная карта QuaD

Критерии оценки	**Вес критерия**	**Баллы**	**Максимальный балл**	**Относительное значение** (3/4)	**Средневзвешенное значение** (5x2)
1	2	3	4	5	
Показатели оценки качества разработки					
1. Удобство в заполнении методик	0,1	70	100	0,7	0,07
2. Продолжительность обработки массива данных	0,2	90	100	0,9	0,18
3. Качество полученных	0,1	60	100	0,6	0,06

результатов					
4. Возможности технологии	0,05	90	100	0,9	0,045
5. Простота алгоритма	0,1	100	100	1	0,1
6. Пути внедрения	0,05	80	100	0,8	0,04
Показатели оценки коммерческого потенциала разработки					
1. Конкурентоспособность технологии	0,1	90	100	0,9	0,09
2. Цена технологии	0,15	95	100	0,95	0,1425
3. Срок внедрения	0,1	90	100	0,9	0,09
4. Наличие сертификации разработки	0,05	100	100	1	0,05
Итого	**1**				**0,8675**

По итогам анализа можно отметить, что разработка является перспективной и в нее стоит инвестировать.

4.1.4 SWOT-анализ

SWOT – Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы) – представляет собой комплексный анализ научно-исследовательского проекта. SWOT-анализ применяют для исследования внешней и внутренней среды проекта.

Первый этап заключается в описании сильных и слабых сторон проекта, в выявлении возможностей и угроз для реализации проекта, которые проявились или могут появиться в его внешней среде.

Сильные стороны – это ресурсы или возможности, которыми располагает руководство проекта и которые могут быть эффективно использованы для достижения поставленных целей.

Слабые стороны – это то, что плохо получается в рамках проекта или где он располагает недостаточными возможностями или ресурсами по сравнению с конкурентами.

Возможности включают в себя любую предпочтительную ситуацию в настоящем или будущем, возникающую в условиях окружающей среды проекта.

Угроза представляет собой любую нежелательную ситуацию, тенденцию или изменение в условиях окружающей среды проекта, которые имеют разрушительный или угрожающий характер для его конкурентоспособности в настоящем или будущем.

Результаты SWOT-анализа представлены на таблице 4.

Таблица 8. SWOT-анализ

	Сильные стороны: С1. Простота алгоритма, заложенного в основе технологии. С2. Простая в заполнении методика. С3. Низкие цены на технологию.	**Слабые стороны:** Сл1. Неквалифицированные пользователи. Сл2. Высокая скорость развития конкурентных технологий.
Возможности: В1. Расширение географии внедрения технологии. В2. Осваивание новых отраслей. В3. Повышение квалификации сотрудников.	1.Внедрение технологии на территории всей страны за счет низкой стоимости. 2. Развитие охватываемых сфер общества благодаря быстрому освоению простого алгоритма технологии.	1. Разработать систему обучения пользователей. 2. Своевременное развитие технологии в соответствии с возникающими новшествами.
Угрозы: У1. Ужесточение конкуренции У2. Выход новых конкурентов на рынок У3. Экономическая нестабильность.	1. Поддерживать существующую ценовую политику для повышения конкурентоспособности. 2. Разработать систему скидок для многоцелевого использования технологии	1.Снижать издержки и оптимизировать ресурсы. 2. Стараться развивать технологию быстрее конкурентов.

Таким образом, в результате SWOT-анализа были выявлены слабые и сильные стороны, а также возможные варианты повышения эффективности и минимизации угроз.

4.2 Планирование проектных работ

4.2.1 Структура работ в рамках проекта

Планирование комплекса предполагаемых работ осуществлено в следующем порядке:

- определение структуры работ проекта;
- определение участников каждой работы;
- установление продолжительности работ;
- построение графика проведения проектной работы.

Для выполнения технического задания была сформирована рабочая группа. По каждому виду запланированных работ установлена соответствующая должность исполнителей.

В данном разделе составлен перечень этапов и работ проекта, а также произведено распределение исполнителей по видам работ (табл. 5).

Таблица 9. Перечень этапов, работ и распределение исполнителей

Основные этапы	**№ раб**	**Содержание работ**	**Должность исполнителя**
Обсуждение идеи технологии	1	Поиск идеи для создания технологии	Руководитель, студент
Анализ исходных данных	2	Предоставление массива многомерных данных	Руководитель
	3	Поиск методик, по которым данные были собраны	Студент
	4	Описание найденных методик	Студент
	5	Выборка данных, необходимых для создания технологии	Руководитель, студент
	6	Выбор инструментария для исследования	Руководитель, студент
Разработка информационной технологии	7	Разработка алгоритма	Руководитель, студент
	8	Анализ и удаление выбросов	Студент
	9	Инициализация и визуализация пропущенных значений	Студент
	10	Восстановление пропущенных данных	Студент
	11	Проверка качества восстановления данных	Руководитель
	12	Сокращение признакового пространства с использованием факторного анализа	Студент

	13	Разбивка данных на 4 группы, используя кластерный анализ	Студент
	14	Интерпретация полученных результатов, сокращение числа используемых методик	Руководитель, студент
Изложение выполненной работы в пояснительной записке	15	Создание отчета по проделанной работе	Студент

4.2.2 Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников проекта.

Трудоемкость выполнения проекта оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов. Для определения ожидаемого (среднего) значения трудоемкости $t_{ожi}$ используется следующая формула:

$$t_{\text{ожі}} = \frac{3t_{\min i} + 2t_{\max i}}{5}, \qquad (3)$$

где $t_{ожi}$ – ожидаемая трудоемкость выполнения *i*-ой работы чел.-дн.;

$t_{\min i}$ – минимально возможная трудоемкость выполнения заданной *i*-ой работы (оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств), чел.-дн.;

$t_{\max i}$ – максимально возможная трудоемкость выполнения заданной *i*-ой работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн.

Расчеты $t_{ожi}$ занесены в таблицу 6.

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях $T_р$, учитывающая параллельность выполнения работ несколькими исполнителями. Такое вычисление необходимо для обоснованного расчета заработной платы, так

как удельный вес зарплаты в общей сметной стоимости проекта составляет около 65 %.

$$T_{\text{р}_i} = \frac{t_{\text{ожі}}}{\text{Ч}_i}, \tag{4}$$

где $T_{\text{р}i}$ – продолжительность одной работы, раб. дн.;

$t_{\text{ож}i}$ – ожидаемая трудоемкость выполнения одной работы, чел.-дн.

Ч_i – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Расчеты продолжительности работ представлены также в таблице 6.

4.2.3 Разработка графика проведения проекта

Наиболее удобным и наглядным способом отслеживания выполнения проектной работы является диаграмма Ганта.

Диаграмма Ганта – горизонтальный ленточный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{\text{к}i} = T_{\text{р}i} \cdot k_{\text{кал}}, \tag{5}$$

где $T_{\text{к}i}$– продолжительность выполнения i-й работы в календарных днях;

$T_{\text{р}i}$ – продолжительность выполнения i-й работы в рабочих днях;

$k_{\text{кал}}$ – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}} = \frac{366}{247} = 1{,}4818, \tag{6}$$

где $T_{\text{кал}}$ – количество календарных дней в году;

$T_{\text{вых}}$ – количество выходных дней в году;

$T_{\text{пр}}$ – количество праздничных дней в году.

Тогда длительность каждого из этапов работ в календарных днях будет равна $T_{\text{к}i} = T_{\text{р}i} \cdot k_{\text{кал}} = T_{\text{р}i} * 1{,}4818$.

Все рассчитанные значения сведены в таблицу 6.

Таблица 10. Временные показатели проведения научного исследования

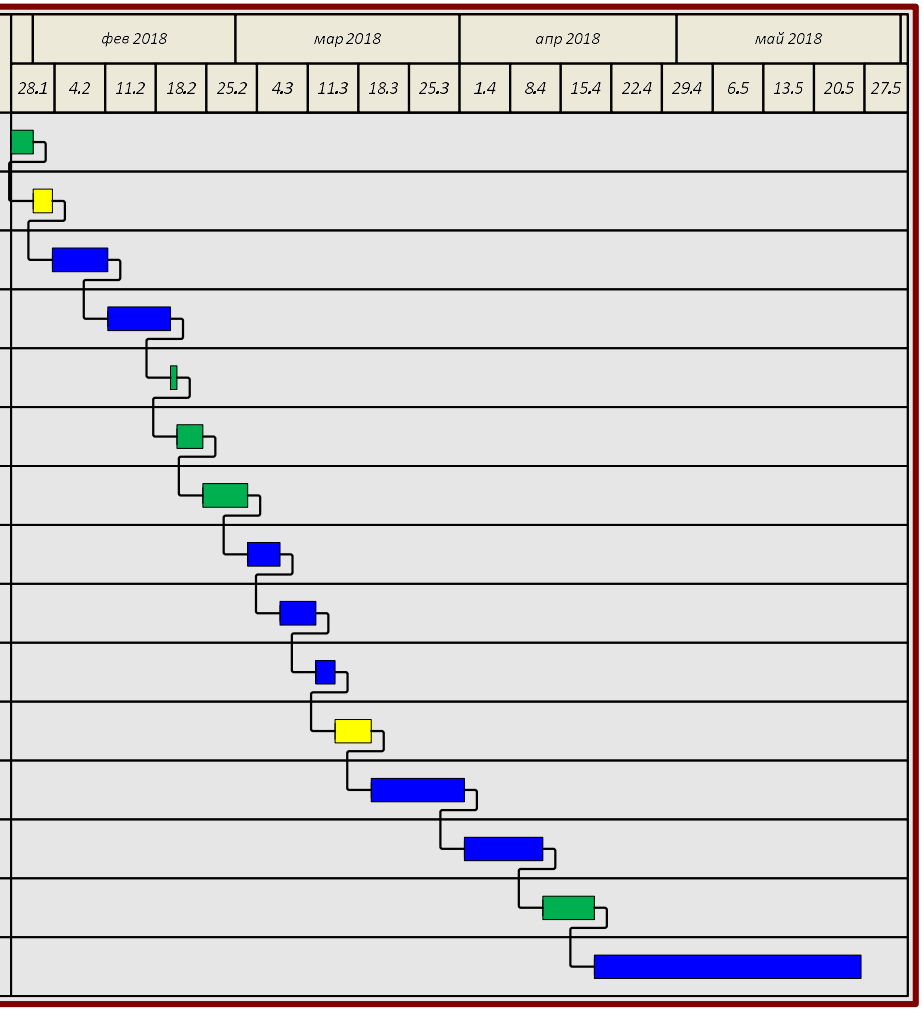
Название работы	**Трудоёмкость работ**			**Исполнители**	**Длительность работ в рабочих днях** $T_{\text{р}i}$	**Длительность работ в календарных днях** $T_{\text{к}i}$
	t_{min}, чел-дни	t_{max}, чел-дни	$t_{ожi}$ чел-дни			
Поиск идеи для создания технологии	3	7	4,6	Руководитель, студент	2,3	3,4
Предоставление массива многомерных данных	1	2	1,4	Руководитель	1,4	2,1
Поиск методик, по которым данные были собраны	5	7	5,8	Студент	5,8	8,6
Описание найденных методик	5	9	6,6	Студент	6,6	9,8
Выборка данных, необходимых для создания технологии	1	2	1,4	Руководитель, студент	0,7	1,0
Выбор инструментария для исследования	3	6	4,2	Руководитель, студент	2,1	3,1
Разработка алгоритма	7	10	8,2	Руководитель, студент	4,1	6,1
Анализ и удаление выбросов	2	4	2,8	Студент	2,8	4,1
Инициализация и визуализация пропущенных значений	3	5	3,8	Студент	3,8	5,6
Восстановление пропущенных данных	1	2	1,4	Студент	1,4	2,1
Проверка качества восстановления данных	3	5	3,8	Руководитель	3,8	5,6

Сокращение признакового пространства с использованием факторного анализа	7	14	9,8	Студент	9,8	14,5
Разбивка данных на 4 группы, используя кластерный анализ	7	10	8,2	Студент	8,2	12,2
Интерпретация полученных результатов, сокращение числа используемых методик	8	12	9,6	Руководитель, студент	4,8	7,1
Создание отчета по проделанной работе	20	40	28	Студент	28	41,5

На основе табл. 6 построен календарный план-график для максимального по длительности исполнения работ в рамках выполняемого проекта. В табл. 7 разбивка по месяцам и неделям за период времени дипломирования. Синим цветом показаны задачи, исполнителем которых являлся только студент, желтым – только руководитель, а зеленым – совместно решаемые задачи.

Таблица 11. Календарный план-график проведения проектной работы

Идентификатор	Название задачи	Начало	Окончание	Длительность
1	Поиск идеи для создания технологии	29.01.2018	01.02.2018	3д ,2ч
2	Предоставление массива многомерных данных	01.02.2018	03.02.2018	2д 6ч
3	Поиск методик, по которым данные были собраны	03.02.2018	11.02.2018	7д 5,8ч
4	Описание найденных методик	11.02.2018	20.02.2018	8д 6,4ч
5	Выборка данных, необходимых для создания технологии	20.02.2018	20.02.2018	8ч
6	Выбор инструментария для исследования	20.02.2018	24.02.2018	3д 5ч
7	Разработка алгоритма	24.02.2018	02.03.2018	6д 2ч
8	Анализ и удаление выбросов	02.03.2018	07.03.2018	4д 4ч
9	Инициализация и визуализация пропущенных значений	07.03.2018	12.03.2018	4д 8,8ч
10	Восстановление пропущенных данных	12.03.2018	14.03.2018	2д 6ч
11	Проверка качества восстановления данных	14.03.2018	19.03.2018	4д 8,8ч
12	Сокращение признакового пространства с использованием факторного анализа	19.03.2018	01.04.2018	12д 8ч
13	Разбивка данных на 4 группы, используя кластерный анализ	01.04.2018	12.04.2018	10д 7,6ч
14	Интерпретация полученных результатов, сокращение числа используемых методик	12.04.2018	19.04.2018	7д 1ч
15	Интерпретация полученных результатов, сокращение числа используемых методик	19.04.2018	26.05.2018	36д 8ч



4.2.4 Бюджет научно-технического исследования (НТИ)

При планировании бюджета НТИ должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. В процессе формирования бюджета НТИ используется следующая группировка затрат по статьям:

- материальные затраты НТИ;
- затраты на специальное оборудование для научных (экспериментальных) работ;
- основная заработная плата исполнителей темы;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- затраты научные и производственные командировки;
- контрагентные расходы;
- накладные расходы.

4.2.4.1 Расчет материальных затрат НТИ

Данная статья включает стоимость всех материалов, используемых при разработке проекта:

- приобретаемые со стороны сырье и материалы, необходимые для создания научно-технической продукции;

- покупные материалы, используемые в процессе создания научно-технической продукции для обеспечения нормального технологического процесса и для упаковки продукции или расходуемых на другие производственные и хозяйственные нужды (проведение испытаний, контроль, содержание, ремонт и эксплуатация оборудования, зданий, сооружений, других основных средств и прочее), а также запасные части для ремонта оборудования, износа инструментов, приспособлений, инвентаря, приборов, лабораторного оборудования и других средств труда, не относимых к основным средствам, износ спецодежды и других малоценных и быстроизнашивающихся предметов;

- покупные комплектующие изделия и полуфабрикаты, подвергающиеся в дальнейшем монтажу или дополнительной обработке;

- сырье и материалы, покупные комплектующие изделия и полуфабрикаты, используемые в качестве объектов исследований (испытаний) и для эксплуатации, технического обслуживания и ремонта изделий – объектов испытаний (исследований);

В материальные затраты, помимо вышеуказанных, включаются дополнительно затраты на канцелярские принадлежности, диски, картриджи и т.п. Однако их учет ведется в данной статье только в том случае, если в научной организации их не включают в расходы на использование оборудования или накладные расходы. В первом случае на них определяются соответствующие нормы расхода от установленной базы. Во втором случае их величина учитывается как некая доля в коэффициенте накладных расходов.

Расчет материальных затрат осуществляется по следующей формуле:

$$З_{м} = (1 + k_{T}) \cdot \sum_{i=1}^{m} Ц_{i} \cdot N_{рас\,xi} \ , \qquad (7)$$

где m – количество видов материальных ресурсов, потребляемых при выполнении научного исследования;

$N_{расхi}$ – количество материальных ресурсов i-го вида, планируемых к использованию при выполнении научного исследования (шт., кг, м, м^2 и т.д.);

$Ц_i$ – цена приобретения единицы i-го вида потребляемых материальных ресурсов (руб./шт., руб./кг, руб./м, руб./м^2 и т.д.);

k_T – коэффициент, учитывающий транспортно-заготовительные расходы.

Значения цен на материальные ресурсы могут быть установлены по данным, размещенным на соответствующих сайтах в Интернете предприятиями-изготовителями (либо организациями-поставщиками).

Величина коэффициента (k_T), отражающего соотношение затрат по доставке материальных ресурсов и цен на их приобретение, зависит от условий договоров поставки, видов материальных ресурсов, территориальной удаленности поставщиков и т.д. Транспортные расходы принимаются в

пределах 15-25% от стоимости материалов. Материальные затраты, необходимые для данной разработки, заносятся в таблицу 8.

Таблица 12. Материальные затраты

Наименование	Единица измерения	Количество	Цена за ед., руб.	Затраты на материалы, (**Зм**), руб.
Ноутбук	шт.	2	32 000	64 000
Компьютерный стол	шт.	2	4 500	9 000
Офисный стул	шт.	2	1 500	3 000
Компьютерная мышь	шт.	2	750	1 500
Итого				**77 500**

4.2.4.2 Основная заработная плата исполнителей темы

В настоящую статью включается основная заработная плата научных и инженерно-технических работников, рабочих макетных мастерских и опытных производств, непосредственно участвующих в выполнении работ по данной теме. Величина расходов по заработной плате определяется исходя из трудоемкости выполняемых работ и действующей системы окладов и тарифных ставок. В состав основной заработной платы включается премия, выплачиваемая ежемесячно из фонда заработной платы в размере 20 –30 % от тарифа или оклада. Расчет основной заработной платы сводится в табл. 9.

Статья включает основную заработную плату работников, непосредственно занятых выполнением НТИ, (включая премии, доплаты) и дополнительную заработную плату:

$$З_{зп} = З_{осн} + З_{доп}, \quad (8)$$

где $З_{осн}$ – основная заработная плата;

$З_{доп}$ – дополнительная заработная плата (12-20 % от $З_{осн}$).

Основная заработная плата ($З_{осн}$) руководителя (программиста) от предприятия (при наличии руководителя от предприятия) рассчитывается по следующей формуле:

$$З_{осн} = З_{дн} \cdot T_p, \quad (9)$$

где $З_{осн}$ – основная заработная плата одного работника;

$Т_р$ – продолжительность работ, выполняемых научно-техническим работником, раб. дн. (табл. 8);

$З_{дн}$ – среднедневная заработная плата работника, руб.

Таблица 13. Расчет основной заработной платы

№ п/п	Наименование этапов	Исполнители по категориям	Трудоемкость, чел.-дн.	Заработная плата, приходящаяся на один чел.-дн., тыс. руб.	Всего заработная плата по тарифу (окладам), тыс. руб.
1	Поиск идеи для создания технологии	Руководитель, студент	4,6	1,50	6,90
2	Предоставление массива многомерных данных	Руководитель	1,4	1,5	2,1
3	Поиск методик, по которым данные были собраны	Студент	5,8	0	0
4	Описание найденных методик	Студент	6,6	0	0
5	Выборка данных, необходимых для создания технологии	Руководитель, студент	1,4	1,50	2,10
6	Выбор инструментария для исследования	Руководитель, студент	4,2	1,50	6,30
7	Разработка алгоритма	Руководитель, студент	8,2	1,50	12,30
8	Анализ и удаление выбросов	Студент	2,8	0	0
9	Инициализация и визуализация пропущенных значений	Студент	3,8	0	0
10	Восстановление пропущенных данных	Студент	1,4	0	0
11	Проверка качества восстановления	Руководитель	3,8	1,5	5,70

	данных				
12	Сокращение признакового пространства с использованием факторного анализа	Студент	9,8	0	0
13	Разбивка данных на 4 группы, используя кластерный анализ	Студент	8,2	0	0
14	Интерпретация полученных результатов, сокращение числа используемых методик	Руководитель, студент	9,6	1,5 0	14,4 0
15	Создание отчета по проделанной работе	Студент	28	0	0
Итого:					49,8

Среднедневная заработная плата рассчитывается по формуле:

$$З_{дн} = \frac{З_м \cdot М}{F_д}, \qquad (10)$$

где $З_м$ – месячный должностной оклад работника, руб.;

М – количество месяцев работы без отпуска в течение года:

при отпуске в 24 раб. дня М =11,2 месяца, 5-дневная неделя;

при отпуске в 48 раб. дней М=10,4 месяца, 6-дневная неделя;

$F_д$ – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн. (табл. 14).

Таблица 14. Баланс рабочего времени

Показатели рабочего времени	**Руководитель**	**Студент**
Календарное число дней	33	94,4
Количество нерабочих дней - выходные дни - праздничные дни	4	11
Потери рабочего времени - отпуск - невыходы по болезни	0	0

Действительный годовой фонд рабочего времени	260	260

Месячный должностной оклад работника:

$$З_м = З_{тс} \cdot (1 + k_{пр} + k_д) \cdot k_р, \quad (11)$$

где $З_{тс}$ – заработная плата по тарифной ставке, руб.;

$k_{пр}$ – премиальный коэффициент, равный 0,3 (т.е. 30% от $З_{тс}$);

$k_д$ – коэффициент доплат и надбавок составляет примерно 0,2 – 0,5 (в НИИ и на промышленных предприятиях – за расширение сфер обслуживания, за профессиональное мастерство, за вредные условия: 15-20 % от $З_{тс}$);

$k_р$ – районный коэффициент, равный 1,3 (для Томска).

Тарифная заработная плата $З_{тс}$ находится из произведения тарифной ставки работника 1-го разряда $Т_{сi}$ = 600 руб. на тарифный коэффициент $k_т$ и учитывается по единой для бюджетных организации тарифной сетке. Для предприятий, не относящихся к бюджетной сфере, тарифная заработная плата (оклад) рассчитывается по тарифной сетке, принятой на данном предприятии. Расчёт основной заработной платы приведён в табл. 15.

Таблица 15. Расчёт основной заработной платы

Исполнители	**$З_{тс}$, руб.**	**$k_{пр}$**	**$k_д$**	**$k_р$**	**$З_м$, руб**	**$З_{дн}$, руб.**	**$Т_р$, раб. дн.**	**$З_{осн}$, руб.**
Руководитель	22 052	0,3	0,2	1,3	43 001,4	1 954,61	29	56 683,69
							Итого $З_{осн}$	56 683,69

4.2.4.3 Дополнительная заработная плата исполнителей темы

Затраты по дополнительной заработной плате исполнителей темы учитывают величину предусмотренных Трудовым кодексом РФ доплат за отклонение от нормальных условий труда, а также выплат, связанных с обеспечением гарантий и компенсаций (при исполнении государственных и общественных обязанностей, при совмещении работы с обучением, при предоставлении ежегодного оплачиваемого отпуска и т.д.).

Расчет дополнительной заработной платы ведется по следующей формуле:

$$З_{доп} = k_{доп} \cdot З_{осн} \quad (12)$$

где $k_{доп}$ – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15).

Таблица 16. Расчёт дополнительной заработной платы

Исполнители	**$З_{осн}$, руб.**	**$k_{доп}$**	**$З_{доп}$, руб.**
Руководитель	56 683,69	0,15	8 502,55
Итого $З_{доп}$			8 502,55

4.2.4.4 Отчисления во внебюджетные фонды (страховые отчисления)

В данной статье расходов отражаются обязательные отчисления по установленным законодательством Российской Федерации нормам органам государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников.

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$З_{внеб} = k_{внеб} \cdot (З_{осн} + З_{доп}), \quad (13)$$

где $k_{внеб}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

На 2016 г. в соответствии с Федерального закона от 24.07.2009 №212-ФЗ установлен размер страховых взносов равный 30,2%.

Отчисления во внебюджетные фонды представлены в таблице 17.

Таблица 17. Отчисления во внебюджетные фонды

Исполнитель	**Основная заработная плата, руб.**	**Дополнительная заработная плата, руб.**
Руководитель проекта	56 683,69	8 502,55
Коэффициент отчислений во внебюджетные фонды	0,302	
Итого:	**19 686,24**	

4.2.4.5 Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии, почтовые и телеграфные

расходы, размножение материалов и т.д. Их величина определяется по следующей формуле:

$$З_{\text{накл}} = (\text{сумма статей}\, 1 \div 4) \cdot k_{\text{нр}}, \qquad (14)$$

где $k_{\text{нр}}$ – коэффициент, учитывающий накладные расходы.

Величина коэффициента накладных расходов 16%.

Таким образом, $З_{\text{накл}} = 162372,48 \cdot 0,16 = 25979,60$

4.2.4.6 Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно-исследовательской работы (темы) является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции.

Определение бюджета затрат на научно-исследовательский проект приведено в табл. 18.

Таблица 18. Расчет бюджета затрат НТИ

Наименование статьи	Сумма, руб.	Примечание
1. Материальные затраты НТИ	77 500	Пункт 2.4.1
2. Затраты по основной заработной плате исполнителей темы	56 683,69	Пункт 2.4.2
3. Затраты по дополнительной заработной плате исполнителей темы	8 502,55	Пункт 2.4.3
4. Отчисления во внебюджетные фонды	19 686,24	Пункт 2.4.4
5. Накладные расходы	25 979,60	16 % от суммы ст. 1-4
6. Бюджет затрат НТИ	188 352,08	Сумма ст. 1- 5

4.3 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с

определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный финансовый показатель разработки определяется как:

$$I_{\text{финр}} = \frac{\Phi_{\text{р}}}{\Phi_{\text{max}}}, \quad (15)$$

где $I_{\text{финр}}$ – интегральный финансовый показатель разработки;

$\Phi_{\text{р}}$ – стоимость исполнения;

Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта (в т.ч. аналоги).

$$I_{\text{финр}} = \frac{188\,352{,}08}{250\,000} = 0{,}75$$

Полученная величина интегрального финансового показателя разработки отражает соответствующее численное удешевление стоимости разработки в разах.

Интегральный показатель ресурсоэффективности исполнения объекта исследования можно определить следующим образом:

$$\mathrm{I}_{\text{р}} = \sum a \cdot b \ , \quad (16)$$

где I_p – интегральный показатель ресурсоэффективности;

a – весовой коэффициент;

b – бальная оценка, устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Расчет интегрального показателя ресурсоэффективности приведен в таблице 19.

Таблица 19. Сравнительная оценка характеристик вариантов исполнения проекта

Критерии / Объект исследования	Весовой коэффициент параметра	Оценка выполнения
1. Способствует росту производительности труда пользователя	0,15	3
2. Удобство в эксплуатации (соответствует требованиям потребителей)	0,25	4
3. Надежность	0,2	4
4. Экономия времени	0,3	5
5. Простота исполнения	0,1	5
ИТОГО	1	

$$I_p = 3 \cdot 0{,}15 + 4 \cdot 0{,}25 + 4 \cdot 0{,}2 + 5 \cdot 0{,}3 + 5 \cdot 0{,}1 = 4{,}25$$

Интегральный показатель эффективности исполнения разработки ($I_{исп.}$) определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{исп.} = \frac{I_p}{I_{финр}} = \frac{4{,}25}{0{,}75} = 5{,}67 \quad (17)$$

Полученное значение интегрального показателя эффективности исполнения разработки превысил максимальный балл в системе оценивания. Это говорит о том, что результат работы можно считать положительным, так как оценка интегрального показателя ресурсоэффективности близка к максимальной, при этом стоимость разработки ниже, чем у ряда аналогов, рассмотренных при анализе конкурентных решений.

В целом данные, полученные при анализе конкурентных решений и оценочной карты Quad, позволяют сделать вывод, что разработка программного продукта является перспективной и привлекательной для инвесторов. Продукт имеет множество преимуществ перед рассмотренными конкурентными решениями, в особенности по таким критериям, как удобство в эксплуатации, функциональные возможности и цена.

SWOT-анализ позволил выявить слабые и сильные стороны, позволяющие повысить эффективность и сократить угрозы, что, в свою очередь, будет способствовать реализации планов по расширению направлений развития.

Также была распланирована структура работ проекта и определены ответственные должности для их выполнения. В соответствии с назначенными работами была рассчитана их трудоемкость и составлен график работ (диаграмма Ганта). Общая длительность проектирования и разработки программного продукта составила 127 дней.

Общий бюджет НТИ составил 188 352,08 рублей. Он включает в себя затраты на основную и дополнительную заработную плату работников, материальные затраты, отчисления на внебюджетные фонды и накладные расходы.

Глава 5. Социальная ответственность

В процессе трудовой деятельности на сотрудника офиса могут оказывать воздействие различного рода производственные факторы. Для их предупреждения и сохранения здоровья работника предусматривается ряд мер по обеспечению безопасности трудовой деятельности.

В данном разделе рассматривается анализ вредных и опасных факторов труда, определяются необходимые меры защиты от них, оцениваются условия труда и предоставляются рекомендации по их оптимизации.

Как правило, офисные работники сталкиваются с повышенным уровнем шума, нарушением температурного режима, недостаточной освещенностью и т.д. Важную роль играют и психофизические факторы: зрительное, слуховое, умственное перенапряжение, монотонность труда и т.д.

Разработка технологии, описанной в данной работе, предполагает работу с информацией, проводимую за персональным компьютером в учебной аудитории №204 Кибернетического центра Национального исследовательского Томского политехнического университета.

Характеристика помещения:

- ширина рабочего помещения 6 м, длина – 6 м, высота – 2,8 м;
- площадь – 36 м^2;
- объём помещения - 100,8 м^3;
- имеется кондиционер, а также естественная вентиляция: двери, окна;
- искусственное освещение;
- естественное освещение.

В данном помещении оборудовано десять рабочих мест, но одновременно в работе обычно задействованы 3-4 человека. Следовательно, в среднем на одного сотрудника приходится не менее 25 м^3 объема помещения и не менее 9 м^2 площади, что удовлетворяет требованиям санитарных норм, согласно которым для одного работника должны быть предусмотрены площадь

величиной не менее 6 м^2 и объем не менее 24 м^3, с учетом максимального числа одновременно работающих в смену.

5.1 Производственная безопасность

Опасные и вредные производственные факторы подразделяются на 4 группы по оказываемому влиянию на человека: физические, химические, биологические и психофизиологические. Так как на состояние офисных работников (программистов) химические и биологические факторы не оказывают существенного влияния, то основное внимание будет уделено физическим и психофизиологическим факторам.

Для представления всех вредных и опасных факторов необходимо классифицировать их в соответствии с нормативными документами.

Таблица 20. Классификация вредных и опасных факторов

Наименование видов работ и параметров производственного процесса	Факторы (ГОСТ 12.0.003-74 ССБТ)		Нормативные документы
	Вредные	Опасные	
1	2	3	4
Работа с компьютером и орг. техникой	1. Повышенная или пониженная влажность воздуха 2. Повышенная (пониженная) температура воздуха 3. Повышенный уровень шума 4. Повышенный уровень электромагнитных излучений 5. Недостаточная освещенность рабочего места 6. Эмоциональные перегрузки 7. Умственное перенапряжение 8. Монотонность труда	1. Опасность поражения электрическим током 2. Вероятность возникновения короткого замыкания 3. Статическое электричество	1. ГОСТ 12.0.003-74 2. СанПиН 2.2.4.548-96 3. ГОСТ 12.1.006–84 4. СанПиН 2.2.1/2.1.1.1278-03 5. СанПиН 2.2.2/2.4.1340-03 6. СНиП 2.04.05-91

5.1.1 Анализ выявленных вредных факторов при разработке и эксплуатации проектируемого решения

5.1.1.1 Микроклимат рабочего помещения

Гигиенические нормативы на параметры микроклимата в рабочей зоне даны в ГОСТ 12.1.005 — 88.

Микроклимат в рабочей зоне определяется действующими на организм человека сочетаниями влажности, температуры воздуха и окружающих поверхностей и скорости движения воздуха.

Мероприятия по доведению микроклиматических показателей до нормативных значений включаются в комплексные планы предприятий по охране труда. Для создания благоприятных условий работы, соответствующих физиологическим потребностям человеческого организма, санитарные нормы устанавливают оптимальные и допустимые метеорологические условия в рабочей зоне помещения.

Согласно СанПиН 2.2.4.548-96 выполняемая работа относится к категории легкая (1б) – интенсивность энергозатрат в пределах 121-150 ккал/час (140-174 Вт), это работы сидя, стоя или связанные с ходьбой с некоторым физическим напряжением.

Таблица 21. Оптимальные величины показателей микроклимата на рабочих местах производственных помещений (СанПиН 2.2.4.548-96)

Период года	Относительная влажность воздуха, %	Скорость движения воздуха, м/с	Температура воздуха, ^{0}C	Температура поверхностей, ^{0}C
Холодный	60-40	0,1	21 - 23	20 – 24
Теплый	60-40	0,1	23-25	22-26

Таблица 22. Допустимые величины показателей микроклимата

Период года	Относительная влажность воздуха, %	Скорость движения воздуха, м/с		Температура воздуха, °С		Температура поверхностей, °С
		для диапазона температур воздуха ниже оптимальных величин, не более	для диапазона температур воздуха выше оптимальных величин, не более	диапазон ниже оптимальных величин	диапазон выше оптимальных величин	
Холодный	15 - 75	0,1	0,2	19,0 - 20,9	23,1 - 24,0	18,0 - 25,0
Теплый	15 - 75	0,1	0,3	20,0 - 21,9	24,1 - 28,0	19,0 - 29,0

В рассматриваемом помещении в холодный период температура поверхностей и температура воздуха составляет 22^0С и 23^0С соответственно, а влажность воздуха 45%; а в теплый период температура поверхностей и температура воздуха – 24^0С и 25^0С соответственно. Сравнивая со значениями из таблицы, отклонений от норм не выявлено.

В данном случае температура воздуха и температура поверхностей составляют 22^0С и 21^0С при относительной влажности 45% в холодный период года; 24^0С и 23^0С при относительной влажности воздуха 50% в теплый период года, что соответствует нормам.

5.1.1.2 Производственное освещение

Освещение – получение, использование и распределение световой энергии для обеспечения благоприятных условий видения предметов и объектов. Освещение оказывает влияние на общее самочувствие и настроение, определяет эффективность труда. Нерационально организованное освещение может явиться причиной травматизма: недостаточно освещенные опасные зоны, слепящие источники света и блики от них, резкие тени и пульсации освещенности ухудшают видимость и могут вызвать неадекватное восприятие наблюдаемого объекта. В компьютерных комнатах должно быть как естественное, так и искусственное освещение. Естественное освещение обеспечивается за счет оконных проемов, коэффициент искусственного освещения (КОЕ) которых должен быть не менее 1,2% в местах, где имеется снежный покров и не менее 1,5% на остальной территории. Свет из окна должен падать с левой стороны от пользователя. Естественное освещение в аудитории осуществляется через два оконных проема размером 1 на 1.35 метра в наружной стене. Нормируемые показатели естественного, искусственного и совмещенного освещения в соответствии с СанПиН 2.2.1/2.1.1.1278-03 указаны в таблице 23.

Таблица 23. Нормируемые показатели естественного, искусственного и совмещенного освещения в соответствии с СанПиН 2.2.1/2.1.1.1278-03

Помещения	Рабочая поверхность и плоскость нормирования КЕО и освещенности и высота плоскости над полом, м	Естественное освещение		Совмещенное освещение		Искусственное освещение				
		КЕО e_н, %		КЕО e_н, %		Освещенность, лк				
		При верхнем или комбинированном освещении	При боковом освещении	При верхнем или комбинированном освещении	При боковом освещении	При комбинированном освещении		При общем освещении	Показатель дискомфорта, М, не более	Коэффициент пульсации освещенности, К_п, %, не более
						всего	от общего			
1	2	3	4	5	6	7	8	9	10	11
Кабинеты, рабочие комнаты	Г – 0,8	3,0	1,0	1,8	0,6	400	200	300	40	15
Помещения для работы с дисплеями и видеотерминалами, залы ЭВМ	Г – 0,8 Экран монитора: В – 1,2	3,5 -	1,2 -	2,1 -	0,7 -	500 -	300 -	400 200	15 -	10

Для организации искусственного освещения в помещениях, в которых работают за персональными компьютерами, рекомендуется применять светильники типа ЛПО36. Также допустимо применять светильники прямого света, преимущественно отраженного света типа ЛПО5, ЛПО13, ЛСО4, ЛПО34, ЛПО31 с люминесцентными лампами типа ЛБ. Ещё допускается применение светильников местного освещения с лампами накаливания. Светильники должны располагаться линиями (прямыми или прерывающимися) так, чтобы при различном расположении компьютеров они были параллельны линии зрения пользователя. Защитный угол светильников должен быть не менее 40 градусов.

Для того чтобы производственное освещение в помещении соответствовало всем нормам, нужно не менее двух раз в год мыть стекла и светильники, а также следить за работой светильников и при необходимости менять вышедшие из строя лампы.

Когда естественного освещения недостаточно, необходимо использовать общее искусственное освещение. В качестве основных источников искусственного освещения используются лампы белого и дневного света ЛБ-20 и ЛД-20.

В помещении, в котором проводилась работа, используются рядно расположенные потолочные светильники с люминесцентными лампами. В результате анализа освещенности рабочего места отклонений от норм выявлено не было. Уровень освещенности соответствует нормам в разные периоды светового дня.

Произведем расчет освещения производственного помещения.

Рассматриваемое помещение имеет светлый цвет потолков и стен, серое покрытие пола. Длина помещения (a) – 6 м., ширина (b) – 6 м., высота (h) – 2,8 м. В качестве источника света используются светильники, каждый из которых содержит по n=4 люминесцентные лампы мощностью 18 Вт.; общая яркость светового потока ($Ф$) 1150 Лм.

Помещение предназначено для работы за персональным компьютером, поэтому нормой освещенности (E) для него согласно СНиП 23-05-95 станет 200-300 Лк, рабочая плоскость стола находится на расстоянии (h_1) 0,8 м. над уровнем пола, коэффициент запаса ($k_з$) равняется 1,4, а коэффициенты отражения: для потолка – 0,7; для стен – 0,5; для пола – 0,3.

Сначала находим площадь помещения (S): 6*6=36 м^2.

Далее находим индекс помещения по формуле $\frac{S}{(h-h_1)\cdot(a+b)} = \frac{36}{(2{,}8-0{,}8)(6+6)} = 1{,}5$.

Теперь на основании показателей отражения поверхностей и высчитанного индекса можно из таблицы определить коэффициент использования ($k_{исп}$). В данном случае он равняется 64.

И, наконец, определим необходимое количество светильников $N = \frac{E\cdot S\cdot 100\cdot k_з}{U\cdot n\cdot Ф} = \frac{300\cdot 36\cdot 100\cdot 1{,}4}{64\cdot 4\cdot 1150} = 5{,}14 \approx 6$.

В помещении, в котором проводилась работа, используются рядно расположенные потолочные светильники с люминесцентными лампами. Проведенные расчеты показали, что минимальное число светильников должно быть равно 6. В результате анализа освещенности рабочего места отклонений

от норм выявлено не было. Уровень освещенности соответствует нормам в разные периоды светового дня.

5.1.1.3 Производственные шумы

Шум – это совокупность различных звуков, возникающих в процессе производственной деятельности и несущих неблагоприятное воздействие на организм человека.

В случае постоянного нахождения при шуме более 85 децибел могут наблюдаться нарушения слуха. Также шум может мешать сконцентрироваться и являться фактором стресса, тем самым повышая систолическое кровяное давление; может привести к несчастным случаям, препятствуя получению предупредительных сигналов.

Для аудитории, в которой осуществлялась работа магистранта, основными источниками шума являются расположенные в помещении компьютеры и кондиционер.

Уровни шума для различных категорий рабочих мест служебных помещений регламентирует ГОСТ 12.1.003-83 «ССБТ. Шум. Общие требования безопасности».

Помещения, в которых для работы используют компьютеры не должны соседствовать с помещениями, в которых уровни шума превышают нормируемые значения. В помещениях, которые оборудованы компьютерами, которые являются основным источником шума, уровень шума на рабочем месте должен быть не более 50 дБ.

Рассматриваемая аудитория по уровню производственных шумов не выходит за рамки допустимых значений. Уровень шума менее 50 дБ.

5.1.1.4 Электромагнитные поля

Во время работы за компьютером человек подвергается воздействию электромагнитного и электростатического полей.

Создаваемое персональным компьютером электромагнитное излучение имеет электрическую (Е) и магнитную (Н) составляющие, а также сложный спектральный состав с диапазоном частот от 0 до 1000 МГц.

Основным источником электромагнитных излучений является монитор, в состав которого входит трансформатор высокой частоты строчной развертки. На сегодняшний день ЭЛТ-мониторы потеряли свою популярность. Их вытеснили ЖК-мониторы, уровень электромагнитного излучения которых гораздо меньше.

СанПиН 2.2.4.1191-03 определяет нормы допустимых уровней напряженности электрических полей. Они зависят от времени пребывания человека в контролируемой зоне. Время допустимого пребывания в рабочей зоне в часах рассчитывается по формуле Т=50/Е–2. Если напряженность электрического поля лежит в диапазоне 20–25 кВ/м, то работа не может продолжаться более 10 минут. При напряженности не превышающей 5 кВ/м деятельность людей в рабочей зоне может осуществляться в течение 8 часов.

Еще одним нормативным документом в данной сфере является СанПиН 2.2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы», регламентирующий безопасные уровни излучений.

В таблицах 24-25 представлены предельно-допустимые уровни напряженности на рабочих местах и допустимые уровни электромагнитных полей.

Таблица 24. Предельно-допустимые уровни напряженности на рабочих местах

Время воздействия за рабочий день, мин	Условия воздействия			
	Общее		локальное	
	ПДУ напряженности кА/м	ПДУ магнитной индукции мТл	ПДУ напряженности кА/м	ПДУ магнитной индукции мТл
0 - 10	24	30	40	50
11 - 60	16	20	24	30
61 - 480	8	10	12	15

Таблица 25. *Допустимые уровни электромагнитных полей согласно СанПиН 2.2.4.1340-03*

Наименование параметра	

Напряженность электромагнитного поля на расстоянии 50 см вокруг дисплея до электрической составляющей, В/м, не более: в диапазоне частот 5 Гц – 2 кГц в диапазоне частот 2 – 400 кГц	 25 2,5
Плотность магнитного потока на расстоянии 50 см вокруг дисплея, нТл, не более: в диапазоне частот 5 Гц – 2 кГц в диапазоне частот 2 – 400 кГц	 250 25
Поверхностный электростатический потенциал, В, не более	500

Для снижения уровня излучений проводятся следующие мероприятия:

- сертификация ПК и аттестация рабочих мест;
- применение фильтров и экранов;
- организационно-технические мероприятия;
- применение средств индивидуальной защиты, направленных на экранирование пользователя ПК целиком или отдельных частей его тела;
- употребление профилактических напитков;
- использование других технических средств защиты от электромагнитных излучений.

Уровень напряженности электромагнитного поля в рассматриваемой аудитории не превышает предельно-допустимые значения. Все рабочие машины прошли сертификацию, а рабочие места аттестованы; индивидуальная защита пользователей не требуется.

5.1.1.5 Психофизиологические факторы

Во время длительной работы за компьютером человек также может подвергаться воздействию психофизиологических факторов, таких как эмоциональные перегрузки, умственное перенапряжение, монотонность труда и другие.

Эмоциональные перегрузки вызывают изменения функционального состояния центральной нервной системы, что может негативно отразиться на состоянии организма в целом. Они могут быть вызваны необходимостью

выполнения большого объема работы, конфликтными или стрессовыми ситуациями.

Умственное перенапряжение может наступать вследствие отсутствия необходимого времени на отдых после продолжительной работы, нарушения режима сна или режима питания. Оно может накапливаться и приводить к возникновению заболеваний.

Отличительными признаками монотонной работы служат однообразие рабочих действий, их многократное повторение и небольшая длительность. Таковой является работа за компьютером. В результате работающий теряет интерес к работе, и у него возникает состояние «производственной скуки». Монотонная работа отрицательно сказывается на эффективности производства: ухудшаются экономические показатели, повышается аварийность, травматизм, растет текучесть кадров.

Также вредным фактором производства может служить фиксированная рабочая поза. Она вызывает нарушение кровообращения в нижних конечностях и органах тазовой области, Которое может приводить к профессиональным заболеваниям, например, варикозное расширение вен.

Для снижения эмоциональных перегрузок и умственных перенапряжений предусмотрены перерывы в работе, возможность выбора удобного времени для выполнения работы. Для уменьшения рисков возникновения последствий от фиксированной рабочей позы установленное в аудитории оборудование имеет регулировки: стул регулируется по высоте и наклону спинки, монитор позволяет подобрать наклон под индивидуальные особенности человека.

5.1.2 Анализ выявленных опасных факторов при разработке и эксплуатации проектируемого решения

5.1.2.1 Электробезопасность

Электробезопасность – это система организационных и технических мероприятий, которые обеспечивают защиту человека от вредного и опасного

для его жизни воздействия электрического тока, электромагнитного поля или статического электричества.

Опасное и вредное воздействия на людей электрического тока и электрической дуги проявляются в виде профессиональных заболеваний и электротравм.

Помещение, в котором расположены персональные компьютеры, относится к помещениям без повышенной опасности, потому что отсутствуют следующие факторы:

- высокая температура;

- токопроводящая пыль;

- токопроводящие полы;

- сырость;

- возможность одновременного прикосновения человека к имеющим соединение с землёй металлоконструкциям зданий, технологическим аппаратам и механизмам и металлическим корпусам электрооборудования.

Мероприятия, направленные на предотвращение возможности поражения электрическим током, включают в себя следующее:

- при выполнении монтажных работ необходимо использовать только исправно работающий инструмент, аттестованный службой КИПиА;

- заземление корпусов приборов и инструментов, которое поможет защитить от поражения электрическим током, который может возникнуть между корпусом приборов и инструментом при пробое сетевого напряжения на корпус;

- запрет на выполнение работ на задней панели при включенном сетевом напряжении;

- выполнение работ по устранению неисправностей должно производиться компетентными людьми;

- нужно постоянно наблюдать за исправностью электропроводки и в случае обнаружения неисправностей незамедлительно принимать действия по их устранению.

Перед началом работы необходимо проверить, чтобы не было свешивающихся со стола или висящих под столом проводов электропитания, убедиться в целостности вилки и провода электропитания, в отсутствии видимых повреждений аппаратуры и рабочей мебели, в отсутствии повреждений и наличии заземления приэкранного фильтра.

Токи статического электричества, которые могут возникнуть в процессе работы компьютера на корпусах системного блока, монитора и клавиатуры, могут провоцировать разряды при прикосновении к этим элементам, которые не представляют опасности для человека, но могут привести к поломке компьютера. Для уменьшения величин токов статического электричества применяются нейтрализаторы, увлажнение воздуха (местное и общее), используются покрытия полов с антистатической пропиткой.

5.2 Экологическая безопасность

Охрана окружающей среды заключается в устранении отходов жизнедеятельности человека и бытового мусора. Если персональные компьютеры теряют свою работоспособность, их списывают и отправляют на специализированный склад, на котором уже принимаются меры по утилизации техники и комплектующих. Под хранением отходов понимается их временное размещение в специально отведённых для этого местах или объектах до их утилизации.

По статистике вышедшие из строя люминесцентные лампы являются одним из самых распространенных источников ртутного загрязнения. Помимо стекла и алюминия каждая лампа содержит приблизительно 60 мг ртути, поэтому отработавшие люминесцентные лампы являются опасным источником токсичных веществ.

Утилизация таких ламп заключается в их передаче перерабатывающим предприятиям, которые имеют специальное оборудование для переработки вредных ламп в безвредное сырье – сорбент, которое может являться материалом для других производств.

Согласно Классификатору отходов ДК 005-96, утвержденному приказом Госстандарта № 89 от 29.02.96 г., отработанные люминесцентные лампы относятся к отходам, которые собираются и сортируются отдельно, поэтому их утилизация и хранение должны отвечать определенным требованиям.

5.3 Безопасность в чрезвычайных ситуациях

В рассматриваемом случае на объекте (аудитория) могут возникнуть чрезвычайные ситуации (ЧС) следующего характера:

- техногенные;
- экологические;
- природные.

Для аудитории, в которой проходит написание ВКР, наиболее вероятно возникновение такой ЧС как пожар, который может возникнуть при замыкании электропроводки оборудования, обрыве проводов или же при несоблюдении мер пожарной безопасности.

Пожарная безопасность – комплекс организационных и технических мероприятий, направленных на обеспечение безопасности людей, на предотвращение пожара, ограничение его распространения, а также на создание условий для успешного тушения пожара.

Помещение, в котором выполняется работа по написанию выпускной работы, относится к категории В по пожарной и взрывной опасности.

К противопожарным мероприятиям в помещении относятся следующие:

1) помещение должно быть оборудовано средствами тушения пожара, такими как огнетушители, стенд с противопожарным инвентарем, ящик с песком; средствами связи; электрическая проводка осветительных приборов и электрооборудования должна быть в исправном состоянии.

2) каждый сотрудник должен знать месторасположение средств тушения пожара и средств связи; знать номера телефонов экстренных служб для оповещения о пожаре; уметь использовать средства пожаротушения.

Рассматриваемое помещение оснащено средствами пожаротушения в соответствии с нормами:

1) огнетушитель пенный ОП-10 – 1 шт.;

2) огнетушитель углекислотный ОУ-5 – 1 шт.

В помещении и на этаже присутствуют следующие средства оповещения:

- световая индикация направления движения к выходу в коридорах этажа;
- звуковая индикация, которая представляет собой систему оповещения о пожаре через громкоговоритель;
- пассивные датчики задымленности.

Чтобы минимизировать вероятность возникновения пожара нужно своевременно проводить профилактические работы, направленные на устранение возможных источников возникновения пожара, такие как:

- систематическое наблюдение за состоянием электропроводки;
- выключение питания оборудования при завершении работы и покидании рабочего места;
- периодическое проведение инструктажа по пожаробезопасности для персонала.

Для того чтобы увеличить устойчивость рабочего помещения к возможному возникновению ЧС, необходимо устанавливать системы противопожарной сигнализации, которые будут реагировать на задымленность и другие продукты горения, размещать необходимые огнетушители, обеспечивать помещение и инструктировать рабочих о плане эвакуации, а также назначить лиц, ответственных за данные мероприятия. Два раза в год (в разные периоды – зимой и летом) проводить учебные тревоги для отработки действий при пожаре.

В ходе осмотра рабочего помещения были обнаружены системы сигнализации о наличии пожара или задымленности и системы пожаротушения.

Если же пожар все-таки возник, необходимо эвакуировать персонал из рабочего помещения в соответствии с планом эвакуации. При отсутствии прямых угроз жизни и здоровью необходимо предпринять возможные меры по тушению источника возгорания огнетушителем. В случае потери контроля над пожаром, необходимо немедленно эвакуироваться по плану эвакуации и ждать приезда специалистов соответствующих служб. При возникновении пожара должна сработать система пожаротушения, издав предупредительные сигналы, и передав на пункт пожарной станции сигнал о ЧС, в случае если система не сработала, по каким-либо причинам, необходимо нажать тревожную кнопку или самостоятельно произвести вызов пожарной службы по телефону 101, сообщить место возникновения ЧС и ожидать приезда специалистов.

5.4 Правовые и организационные вопросы обеспечения безопасности

Для обеспечения безопасности при работе определяют следующие требования к организации рабочих мест пользователей:

- рабочее место должно быть организовано с учетом эргономических требований согласно ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования» и ГОСТ 12.2.061-81 «ССБТ. Оборудование производственное. Общие требования безопасности к рабочим местам»;
- конструкция рабочей мебели (компьютерный стол, офисное кресло, подставка для ног) должна предусматривать возможность регулировки в соответствии с индивидуальными особенностями пользователя для создания комфортных условий для выполнения работы. Вокруг ПК должно быть обеспечено свободное пространство в радиусе как минимум 60-120см;
- оригинал-держатель должен быть установлен на уровне экрана.

В соответствии с государственными стандартами и правовыми нормами обеспечения безопасности предусмотрена рациональная организация труда в течение смены, которая предусматривает:

- продолжительность рабочей смены, не превышающей 8 часов;
- длительность обеденного перерыв не меньше 40 минут;
- установление двух регламентируемых перерывов (не меньше 20 минут после 1-2 часов работы и не меньше 30 минут после 2 часов работы).

Обязательно должен быть предусмотрен предварительный медосмотр, который осуществляется при приеме на работу, и периодические медосмотры.

Также перед приемом на работу каждый сотрудник должен пройти инструктаж по технике безопасности, а в дальнейшем с работником должен быть проведен инструктаж по электробезопасности и охране труда.

5.5 Выводы по разделу

В данном разделе были рассмотрены основные аспекты производственной, экологической и техногенной безопасности. В рамках производственной безопасности были изучены микроклимат производственного помещения, который включает в себя анализ освещенности, шума и электромагнитных полей, и психофизиологические факторы; а также выполнен расчет минимального количества светильников, необходимых для обеспечения необходимого уровня освещенности. Экологическая безопасность сводится к утилизации вредных отходов производства. Техногенная безопасность заключается в проведении необходимых мер по предотвращению возникновения чрезвычайных ситуаций.

В результате анализа всех факторов, рассматриваемое помещение является полностью безопасным для работы и соответствует нормативам.

Заключение

В результате исследования исходных данных была разработана информационная технология оценки показателей качества жизни пациентов, которая позволила сократить количество используемых методик до трех. В конечный список вошли тест SF-36 health status survey, шкала базисных убеждений Янов-Бульман и тест жизнестойкости Мадди (в адаптации Леонтьева). Среднее время тестирования сократилось с 48 до 30 минут, то есть на 37,5%.

Был разработан алгоритм обработки данных, заложенный в основу разработанной информационной технологии. Он состоит из нескольких этапов:

- поиск аномальных значений и, по возможности, их удаление;
- восстановление однородности выборки;
- сокращение признакового пространства путем выявления латентных переменных;
- интерпретация полученных результатов.

По результатам исследований были написаны следующие статьи:

1. Былина Т. А. Методы Data Mining в задачах принятия решений / Т. А. Былина ; науч. рук. О. В. Марухина // Молодежь и современные информационные технологии : сборник трудов XV Международной научно-практической конференции студентов, аспирантов и молодых учёных, 04-07 декабря 2017 г., г. Томск. — Томск : Изд-во ТПУ, 2017. — [С. 264-265].
2. Былина Т. А. Применение методов Data Mining в исследовательских задачах / Т. А. Былина ; науч. рук. О. В. Марухина // Информационные технологии в науке, управлении, социальной сфере и медицине : сборник научных трудов IV Международной научной конференции, 5-8 декабря 2017 г., Томск : в 2 ч. — Томск : Изд-во ТПУ, 2017. — Ч. 1. — [С. 361-362].

Список использованных источников

1. Райзберг Б. А., Лозовский Л. Ш., Стародубцева Е. Качество жизни. // Б. Современный экономический словарь. — 2-е изд., испр. М.: ИНФРА-М, 1999. — 479 с.
2. Л. Ф. Ильичёв, П. Н. Федосеев, С. М. Ковалёв и др. Качество жизни. // Философский энциклопедический словарь. — М.: Советская энциклопедия, 1983.
3. Зубец А. Н. Истоки и история экономического роста. — М.: Изд-во "Экономика", 2014. — 463 с.
4. Зубец А.Н., Тарба И.В. Качество жизни в России / Журнал «Финансы» М., № 12, 2013. — с. 68-70
5. Новик А. А., Ионова Т. И. Руководство по исследованию качества жизни в медицине. 2-е издание / Под ред. Л. Шевченко.— М.:ЗАО «ОЛМА Медиа Групп», 2007.— 320с.
6. 36-Item Short Form Survey (SF-36) [Электронный ресурс] — Код доступа: http://www.rand.org/health/surveys_tools/mos/mos_core_36item.html
7. Собчик, Л. Н. Метод цветовых выборов — модификация восьмицветового теста Люшера : практическое руководство. — СПб. : Речь, 2007. — 128 с
8. Люшер М. Цвет вашего характера. — Москва: РИПОЛ КЛАССИК, 1997. — С. 14-15. — 240 с.
9. Падун М.А., Котельникова А.В. Методика исследования базисных убеждений личности. Лаборатории психологии и психотерапии посттравматического стресса. — М.:ИПРАН, 2007.
10. Леонтьев Д.А., Рассказова Е.И. Тест жизнестойкости. Методическое руководство по новой методике психологической диагностики личности с широкой областью применения. Предназначается для профессиональных психологов-исследователей и практиков. - М.: Смысл, 2006.
11. Мандрикова Е.Ю. Разработка опросника самоорганизации деятельности (ОСД). // Психологическая диагностика, 2010, №2. С. 87-111.
12. Былина Т. А. Методы Data Mining в задачах принятия решений / Т. А. Былина; науч. рук. О. В. Марухина // Молодежь и современные информационные технологии : сборник трудов XV Международной научно-практической конференции студентов, аспирантов и молодых учёных, 04-07 декабря 2017 г., г. Томск. — Томск : Изд-во ТПУ, 2017. — [С. 264-265].

13. Зайдель А.Н. Элементарные оценки ошибок измерений. — М.: Наука, 1965.
14. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. [Электронный ресурс] — Код доступа: http://r-analytics.blogspot.com
15. Гусев А. Н., Измайлов Ч. А., Михалевская М. Б. Измерение в психологии. — М.: Смысл, 1997. — 287 с.
16. Факторный, дискриминантный и кластерный анализ / сборник работ под ред. Енюкова И. С. — М.: Финансы и статистика, 1989. — 215 с.
17. Митина О. В., Михайловская И. Б. Факторный анализ для психологов. — М.: Учебно-методический коллектор Психология, 2001. — 169 с.
18. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Классификация и снижение размерности. — М.: Финансы и статистика, 1989. — 607 с.
19. Мандель И. Д. Кластерный анализ. — М.: Финансы и статистика, 1988. — 176 с.
20. Бериков В. С., Лбов Г. С. Современные тенденции в кластерном анализе // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы», 2008. — 26 с.
21. Савельев А. А., Мухарамова С. С., Пилюгин А. Г. Основные понятия языка R. Учебно-методическое пособие. — Казань: Казанский гос. ун-т, 2007. — 29 с.
22. Зарядов И. С. Введение в статистический пакет R: типы переменных, структуры данных, чтение и запись информации, графика. — М.: Издательство Российского университета дружбы народов, 2010 — 207 с.
23. Кузьмина Е.А., Кузьмин А.М. Методы поиска новых идей и решений "Методы менеджмента качества" №1 2003 г.
24. Кузьмина Е.А, Кузьмин А.М. Функционально-стоимостный анализ. Экскурс в историю. "Методы менеджмента качества" №7, 2002 г.
25. Основы функционально-стоимостного анализа: Учебное пособие / Под ред. М.Г. Карпунина и Б.И. Майданчика. — М.: Энергия, 1980. — 175 с.
26. 4Скворцов Ю.В. Организационно-экономические вопросы в дипломном проектировании: Учебное пособие. — М.: Высшая школа, 2006. — 399 с.
27. Сущность методики FAST в области ФСА [Электронный ресурс] — Код доступа:http://humeur.ru/page/sushhnost-metodiki-fast-v-oblasti-fsa.
28. Белов С.В. Безопасность жизнедеятельности и защита окружающей среды (техносферная безопасность) [Электронный ресурс] : учебник для бакалавров / С. В. Белов. — 4-е изд. — Мультимедиа ресурсы (10

директорий; 100 файлов; 740MB). — Москва: Юрайт, 2013. — Код доступа: http://www.lib.tpu.ru/fulltext2/m/2013/FN/fn-2440.pdf
29. Кукин П.П. Безопасность жизнедеятельности. Безопасность технологических процессов и производств. Охрана труда : учебное пособие для вузов— 5-е изд., стер. — М.: Высшая школа, 2009. — 335 с.
30. Беляков Г.И. Охрана труда и техника безопасности [Электронный ресурс] : учебник для прикладного бакалавриата / Г. И. Беляков. — 3-е изд., перераб. и доп. — Мультимедиа ресурсы (10 директорий; 100 файлов; 740MB). — Москва: Юрайт, 2016. — Код доступа: http://www.lib.tpu.ru/fulltext2/m/2015/FN/fn-89.pdf
31. ГОСТ 12.0.003-74. ССБТ. Опасные и вредные производственные факторы. Классификация
32. ГОСТ 12.1.006–84 ССБТ. Электромагнитные поля радиочастот. Общие требования безопасности.
33. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений.
34. СанПиН 2.2.1/2.1.1.1278–03. Гигиенические требования к естественному, искусственному и совмещённому освещению жилых и общественных зданий.
35. ГОСТ 12.1.005-88. ССБТ. Общие санитарно-гигиенические требования к воздуху рабочей зоны
36. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений.
37. СанПиН 2.2.1/2.1.1.1278–03. Гигиенические требования к естественному, искусственному и совмещённому освещению жилых и общественных зданий.
38. ГОСТ 12.1.003–83 ССБТ. Шум. Общие требования безопасности.
39. СанПиН 2.2.4.1191–03. Электромагнитные поля в производственных условиях.
40. СанПиН 2.2.2/2.4.1340–03. Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».
41. ГОСТ 12.2.061-81 ССБТ. Оборудование производственное. Общие требования безопасности к рабочим местам
42. ГОСТ 12.2.032-78. Система стандартов безопасности труда. Рабочее место при выполнении работ сидя. Общие эргономические требования.

Приложение А

Раздел 2

Methods of Initial Data Analysis

Студент:

Группа	ФИО	Подпись	Дата
8КМ61	Былина Татьяна Андреевна		

Консультант отделения ИТ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Чердынцев Евгений Сергеевич	к.т.н.		

Консультант – лингвист отделения ИЯ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Краснова Татьяна Ивановна			

1 Preparation of initial data for analysis

1.1 Analysis of outliers

Outliers are the indicators that are significantly different from the other ones within the selection that is considered.[1]

There are several reasons that can result in outliers:

- measurement error;
- data entry error;
- data interpretation error;
- features of the data nature.

To obtain more accurate analysis results, the outliers must be removed. To detect outliers, the simplest methods based on the midspread distance are used: the outliers are the values beyond the range $\left[\left(x_{25} - 1{,}5 \cdot (x_{75} - x_{25})\right), \left(x_{25} + 1{,}5 \cdot (x_{75} - x_{25})\right)\right]$.[2]

The outliers are analyzed by using box plots. They are also called "box-and-whisker." They show a one-dimensional distribution of probability.

This diagram shows the median, the lower and upper quartiles, the minimum and maximum value of the sample, and the outliers. Some of these boxes can be drawn side by side in order to be visually compared with each other. They can be drawn either horizontally or vertically. Distances between different parts of the box allow you to determine the degree of dispersion and data asymmetry and also to define outliers (Fig. 1).[2]

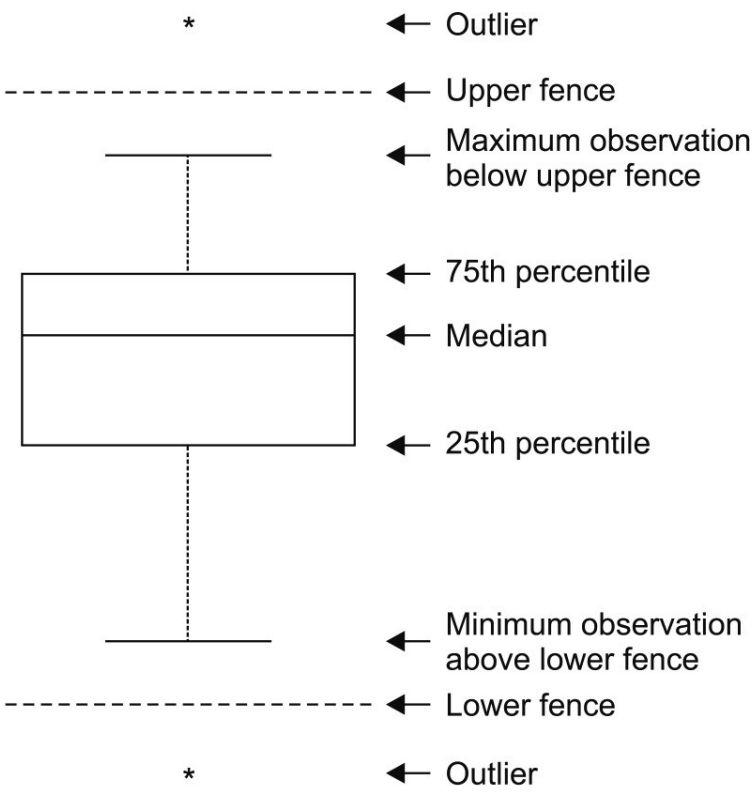


Fig. 1. Boxplot with outliers

1.2 Recovering of Missed Values

Unfortunately, most of the statistical methods assume that during the researches, full-sized matrices, vectors and other information structures of the experiment were obtained. Since data gaps are still widespread in practice, before starting analytical research, it is necessary to bring the tables that are to be processed to a "canonical" form, i.e. either to delete fragments of objects with missed elements, or to replace existing gaps with some reasonable values.

Despite the fact that statistical books sketch out the problem of researching missed data, there are an impressive array of approaches, methodologies and their critical analyzes in this area. In practice, the procedure for "combating passes" usually includes the following steps:

1. identification of missed data;
2. an investigation of the regularities of the appearance of missed values;
3. generating data sets that do not contain gaps.[3]

It is necessary to take into account that identification of missed data is the only unambiguous step. Analysis of why the data is not available depends on your understanding of the processes that reproduce the experimental information. The decision on how to recover missed values will also depend on your assessment of which procedures will lead to the most reliable and accurate results.

The study of the regularities of the presence of missed values is made in order to get an idea of the possible mechanisms for the appearance of missed data and the influence of missed data on the quality of answers to questions of interest to us.

In particular, it is necessary to know what proportion of the data is missing, whether the missed data is concentrated in several variables, or whether they are widely distributed throughout the data set, whether they can be considered random and whether covariance of missed data with each other or with observable data reveals a possible mechanism lying the basis of the missing values.

Answers to these questions will help to determine which statistical methods are best suited for data analysis.

Rational approach. With the so-called rational approach, attempts to replace or restore missed values use mathematical or logical relationships between variables.

To apply a rational approach, creative thinking is usually required, along with proper data management skills. Data recovery can be accurate or approximate.[3]

Complete-case analysis (line delete). In the analysis of complete rows only work with strings without missed values. Many widely used statistical packages by default use listwise / case-wise deletion when working with missed data. Such an approach is so common that many analysts, when conducting regression or dispersion analysis, may not even realize that there is a "problem of missed data" which you need to solve.

With progressive data deletion, it is assumed that the gaps are completely random (that is, full lines are a random selection from the entire data set). Removing all rows with missed data can reduce statistical power by decreasing the selection size.[3]

Multiple imputation (MI) is a way to fill in gaps by re-modeling. Multiple imputation is often used to work with missed data in complex situations. With this approach, several complete data sets are created from the existing data set with missed values (usually three to ten). To replace missed values in derived data sets, Monte Carlo methods are used.

Standard statistical methods are applied to each of the derived data sets, and based on their results, final outcome estimates and confidence intervals are generated that take into account the uncertainty created by missed values.[3]

1.3 Factor Analysis

Factor analysis is used to study the relation between the values of variables. It is assumed that the known variables depend on fewer unknown variables and a random error.[4]

There are two basic concepts of factor analysis: factor – hidden variable and load – the correlation between the original variable and the factor.

The tasks and possibilities of factor analysis

Factor analysis allows solving two important problems of the researcher: to describe the object of measurement comprehensively and at the same time compactly. With the help of factor analysis, it is possible to identify hidden variables that are responsible for the presence of linear statistical correlations between the observed variables. [5]

The main objectives of factor analysis:

- definition of relation between variables, (classification of variables), that is, "objective R-classification";
- reducing the number of variables needed to describe the data. [5]

In the analysis, strongly correlated variables are combined into one factor, as a result a redistribution of the dispersion between the components takes place and the most simple and obvious structure of the factors is obtained. After combining, the correlation of components within each factor among themselves will be higher than their correlation with components from other factors. This procedure also allows you

to identify latent variables, which is especially important in the analysis of social representations and values. For example, by analyzing the scores obtained from several scales, the researcher observes that they are similar to each other and have a high correlation coefficient, the researcher can assume that there is some latent variable by which one can explain the observed similarity of the estimates obtained. Such a latent variable is called a factor. This factor affects the numerous indicators of other variables, which leads us to the possibility and necessity to distinguish it as the most general, higher order. To identify the most significant factors and, as a consequence, the factor structure, it is most justified to apply the principal component method. The essence of this method consists in replacing correlated components with uncorrelated factors. Another important characteristic of the method is the ability to confine itself to the most informative main components and exclude the rest from the analysis, which simplifies the interpretation of the results. The advantages of the principal component method is that it is the only mathematically grounded method of factor analysis. According to a number of researchers, the IGC is not a factor analysis method, since it does not split the variance of indicators into a common and unique one. The main meaning of factor analysis is to isolate from the whole set of variables only a small number of latent independent groups from each other, inside which the variables are related more strongly than variables belonging to different groups.

Factor analysis can be:

- exploratory – it is carried out in the study of a hidden factor structure without the assumption of the number of factors and their loads;
- confirmatory – designed to test hypotheses about the number of factors and their loads. [6]

Conditions for the applying of factor analysis

The practical implementation of factor analysis begins with the verification of its conditions. The mandatory conditions for applying the factor analysis include:

- all signs must be quantitative;
- the number of observations must be at least twice as large as the number of variables;

- the selection must be homogenous;
- the source variables must be symmetrically distributed;
- factor analysis is performed by correlating variables. [5]

1.4 Cluster Analysis

Cluster analysis is a multidimensional statistical procedure that collects data that contains information about the selection of objects, and then sorts objects into relatively homogeneous groups.[7]

The spectrum of applications of cluster analysis is very wide: it is used in archeology, medicine, psychology, chemistry, biology, public administration, philology, anthropology, marketing, sociology, geology and other disciplines. However, the flexibility of this application has led to the emergence of a large number of incompatible terms, methods and approaches that make it difficult to use unambiguously and a consistent interpretation of cluster analysis.

Cluster analysis performs the following main tasks:

- development of a typology or classification;
- research of useful conceptual schemes of grouping of objects;
- generation of hypotheses based on data research;
- test hypotheses or researches to determine whether the types (groups) identified in one way or another are actually present in the available data. [7]

Regardless of the subject of the study, the use of cluster analysis involves the following stages:

- perform a selection for clustering. It is needed to cluster quantitative data only;
- define the set of variables by which objects in the selection will be evaluated, that is, the characteristic space;
- calculating a degree of similarity (or difference) between objects;
- apply the cluster analysis method to create groups of similar objects;
- verify the reliability of the results of the cluster solution. [8]

One can meet the description of two fundamental requirements for data - homogeneity and completeness. Homogeneity requires that all clustered entities be of

the same nature, described by a similar set of characteristics. If the cluster analysis is preceded by factor analysis, then the sample does not need to be "repaired" - the requirements set forth are automatically performed by the factor modeling procedure itself. Otherwise, the sample needs to be adjusted.

Clustering Objectives

- understanding the data by identifying the cluster structure. Splitting a sample into groups of similar objects allows you to simplify further processing of data and decision making by applying to each cluster your analysis method (the "divide and conquer" strategy);
- data compression. If the original selection is excessively large, then we can shorten it, leaving one of the most typical representatives from each cluster;
- novelty detection. Atypical objects are selected which can not be attached to any of the clusters. [7]

In the first case, the number of clusters tries to make less. In the second case, it is more important to ensure a high degree of similarity of objects within each cluster, and there may be as many clusters as possible. In the third case, the most interesting are individual objects that do not fit into any of the clusters.

In all these cases, hierarchical clustering can be used, when large clusters break up into smaller clusters, which in turn are fragmented even smaller, etc. Such tasks are called taxonomy tasks. The result of the taxonomy is a tree-like hierarchical structure. Each object is characterized by enumeration of all clusters to which it belongs, usually from large to small.

Clustering Methods

There is no generally accepted classification of clustering methods, but a number of groups of approaches can be distinguished (some methods can be attributed to several groups at the same time and therefore it is suggested to consider this typing as some approximation to the real classification of clustering methods):

- the probabilistic approach assumes that each object under consideration belongs to one of the k classes. Some authors (for example A.I. Orlov) believe that this group does not belong to clustering at all and contrast it with the name

"discrimination", that is, the choice of assigning objects to one of the known groups (training samples);

- approaches based on artificial intelligence systems: a very conventional group, as there are a lot of methods and methodically they are very different;
- logical approach - the construction of a dendrogram is carried out using a decision tree;
- theoretical-graph approach;
- hierarchical approach presupposes the presence of nested groups (clusters of different orders). Algorithms in turn are divided into agglomerative (unifying) and divisive (separating). By the number of signs, monothetical and polythetic classification methods are sometimes distinguished;
- other methods not included in the previous groups. [7]

1.4.1 The k-means

The k-means method is the most popular method of clustering. It was invented in the 1950s by the mathematician Hugo Steinghaus and almost simultaneously with Stuart Lloyd. Especially popular after McQueen's work.

The algorithm is aimed at minimizing the total quadratic deviation of cluster points from the centers of these clusters:

$$V = \sum_{i=1}^{k} \sum_{x \in S_i} (x - \mu_i)^2,$$

$where\ k - number\ of\ cluster,$

$S_i -\ the\ resulting\ clusters,$

$i = 1, 2, 3, \dots, k,$

$\mu_i -\ the\ centers\ of\ mass\ of\ all\ the\ x\ from\ the\ cluster\ S_i$.[8]

Cluster centers are also called the main points, and the method is called the principal point method and is included in the general theory of principal objects that ensure the best approximation of the data.

The algorithm is a version of the EM algorithm, which is also used to separate the Gaussian mixture. It divides the set of elements of a vector space into a number of clusters k known in advance.

The basic idea is that at each iteration the center of mass is recalculated for each cluster obtained in the previous step, then the vectors are divided into clusters again according to which of the new centers is closer to the selected metric.

The algorithm is completed when no intra-cluster distance changes at some iteration. This happens for a finite number of iterations, since the number of possible partitions of a finite set is finite, and at each step the total quadratic deviation V decreases, so cycling is impossible.

K-means problems:

- the global minimum of the total quadratic deviation V, but only one of the local minima, is not guaranteed;
- the result depends on the choice of the initial centers of the clusters, their optimal choice is unknown;
- the number of clusters must be known in advance.[8]

2 Description of the Tool

2.1 The Scripting Language R

R is a programming language for statistical processing of data and working with graphics. Originally, R was developed by the staff of the statistical department of the University of Auckland Ross Eyhaka and Robert Gentleman; language and environment are supported and developed by the R Foundation.[9]

R is widely used as a statistical software for data analysis and has actually become the standard for statistical programs.

R uses the command-line interface, although several graphical user interfaces are available, such as R Commander, RKWard, RStudio, Weka, Rapid Miner, KNIME, and integration tools in office packages.

2.2 Features of R

R supports a wide range of statistical and numerical methods and has good extensibility with the help of packages. Packages are libraries for the operation of specific functions or special applications. The basic package R includes the main set of packages, and as of the year 2017, more than 11778 packages are available.[9]

Another feature of R is the ability to create high-quality graphics, which can include mathematical symbols.

Advantages of the environment R:

- free and cross-platform;
- a rich arsenal of statistical methods;
- quality vector graphics;
- more than 7000 tested packages;
- flexible in use:
    - allows you to create and edit scripts and packages,
    - interacts with other languages, such as C, Java, and Python,
    - can work with data formats for SAS, SPSS and STATA;
- active community of users and developers;
- regular updates, good documentation and those, support.

The main drawback is a small amount of information in Russian.

2.3 Data Formats

From the statistical point of view, data is divided into types depending on how closely they can be represented using the known metaphor of the numerical line. For example, the age of a person is easy to imagine in this way, except that it can not be negative. The size of the shoes is so much more difficult to present because there is usually no intermediate value between the two adjacent sizes. While there is always something intermediate between any two numbers on a number line. But the dimensions can be at least arranged in increasing or decreasing order. But the sex of a person so to imagine is not at all possible: there are only two values, and "intermediate" simply does not happen. Of course, we can designate a female gender as a unit, and a male as a zero (or a deuce), but they will not carry any numerical information - they can not even be sorted. There are also other special types of data, for example, angles, geographic coordinates, dates, etc., but all of them can somehow be represented by numbers. Thus, the most fundamental difference between data types is whether or not they can be represented using "ordinary" numbers. If it is impossible, then such data is usually called categorial. Statistical laws, and therefore statistical programs, work with such data only if their type is specified in advance. The other types of data in different books are called differently: numeric, counting, ordinal or non-categorical.[9]

So, the main types of data are:

- Number vectors;
- Factors;
- Missed data;
- Matrices;
- Lists.

References

1. Balakrishnan, N., Childs, A. Encyclopedia of Mathematics, Springer Science+Business Media B.V. / Kluwer Academic Publishers, 2001.
2. Maddala, G. S. Introduction to Econometrics – 2nd ed. – New York: MacMillan, 1992 – 88–96.
3. Zairate L.E., Nogueira B.M., Santos T.R.A., Song M.A.J. Techniques for Missing Value Recovering in Imbalanced Databases: Application in a Marketing Database with Massive Missing Data. // International Conference on Systems, Man, and Cybernetics. – 2006.
4. Child D,.The Essentials of Factor Analysis – 3rd ed. – Continuum International. – 2006.
5. Mulaik, S. A. Foundations of Factor Analysis. – Chapman & Hall. – 2010.
6. Fabrigar L.R., Wegener D.T., MacCallum, R.C., Strahan, E.J. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods. – 1999.
7. Owen K.J., James R.F. Function-point cluster analysis. Systematic Zoology – 1973. – 295–301.
8. Bartholomew D.J., Steele F., Galbraith J., Moustaki I. Analysis of Multivariate Social Science Data. Statistics in the Social and Behavioral Sciences Series – 2nd ed. – Taylor & Francis, 2008.
9. Hornik K. R FAQ. The Comprehensive R Archive Network. 2.1 What is R? – 2015.