

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

Институт Кибернетики
Направление подготовки 09.04.01 Информатика и вычислительная техника
Кафедра Информационных систем и технологий

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Статистические методы оценки качества обучения студентов ТПУ

УДК 378.662:658.562:519.2(571.16)

Студент

Группа	ФИО	Подпись	Дата
8ВМ5А	Чернета Дмитрий Сергеевич		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ИСТ	Кацман Юлий Янович	к.т.н., доцент		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. менеджмента	Конотопский Владимир Юрьевич	к.э.н., доцент		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ЭБЖ	Извеков Владимир Николаевич	к.т.н., доцент		

ДОПУСТИТЬ К ЗАЩИТЕ:

Зав. кафедрой	ФИО	Ученая степень, звание	Подпись	Дата
Зав. кафедрой ИСТ	Мальчуков Андрей Николаевич	к.т.н., доцент		

**ЗАПЛАНИРОВАННЫЕ РЕЗУЛЬТАТЫ ПО ОСНОВНОЙ
ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЕ ПОДГОТОВКИ МАГИСТРОВ
09.04.01 «ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА», ИК
ТПУ, ПРОФИЛЬ «КОМПЬЮТЕРНЫЙ АНАЛИЗ И ИНТЕРПРЕТАЦИЯ
ДАННЫХ»**

Планируемые результаты обучения

Код результата в	Результат обучения (выпускник должен быть готов)	Требования ФГОС ВО (ФГОС 3+), критерии АИОР, заинтересованных работодателей и студентов
Общепрофессиональные компетенции		
P1	Воспринимать и самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.	Требования ФГОС 3+ (ОПК-1; ПК 3-6; ОК-4), критерий 5 АИОР (п. 1.1), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
P2	Владеть и применять методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях.	Требования ФГОС 3+ (ОПК-5; ПК-7; ОК-7), критерий 5 АИОР (п. 1.1, 1.2), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
P3	Демонстрировать культуру мышления, способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.	Требования ФГОС 3+ (ОПК-6; ПК-1,2; ОК-1,2), критерий 5 АИОР (п. 1.2), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
P4	Анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности. Владеть, по крайней мере, одним из иностранных языков на уровне социального и профессионального общения, применять	Требования ФГОС 3+ (ОПК-3,4; ПК-11,12; ОК-3), критерий 5 АИОР (п. 1.6, п. 2.2), соответствующий

Код результата в	Результат обучения (выпускник должен быть готов)	Требования ФГОС ВО (ФГОС 3+), критерии АИОР, заинтересованных работодателей и студентов
	специальную лексику и профессиональную терминологию языка.	международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
Профессиональные компетенции		
Р5	Выполнять инновационные инженерные проекты по разработке аппаратных и программных средств автоматизированных систем различного назначения с использованием современных методов проектирования, систем автоматизированного проектирования, передового опыта разработки конкурентно способных изделий.	Требования ФГОС 3+ (ПК-8–12; ОПК-2, ПК-7,6), критерий 5 АИОР (п. 1.3), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
Р6	Планировать и проводить теоретические и экспериментальные исследования в области проектирования аппаратных и программных средств автоматизированных систем с использованием новейших достижений науки и техники, передового отечественного и зарубежного опыта. Критически оценивать полученные данные и делать выводы.	Требования ФГОС 3+ (ПК-1–7; ОПК-6; ОК-4,9), критерий 5 АИОР (п.1.4), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
Р7	Осуществлять авторское сопровождение процессов проектирования, внедрения и эксплуатации аппаратных и программных средств автоматизированных систем различного назначения.	Требования ФГОС 3+ (ПК-13–19; ОПК-5; ОК-8), критерий 5 АИОР (п. 1.5), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.

Код результата в	Результат обучения (выпускник должен быть готов)	Требования ФГОС ВО (ФГОС 3+), критерии АИОР, заинтересованных работодателей и студентов
Общекультурные компетенции		
Р8	Использовать на практике умения и навыки в организации исследовательских, проектных работ и профессиональной эксплуатации современного оборудования и приборов, в управлении коллективом.	Требования ФГОС 3+ (ОК-5,8; ОПК-1,6; ПК-6,7,11,12), критерий 5 АИОР (п. 2.1, п. 2.3, п. 1.5), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
Р9	Осуществлять коммуникации в профессиональной среде и в обществе в целом, активно владеть иностранным языком, разрабатывать документацию, презентовать и защищать результаты инновационной инженерной деятельности, в том числе на иностранном языке.	Требования ФГОС 3+ (ОК-2,9; ОПК-4; ПК-1), критерий 5 АИОР (п. 2.2), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
Р10	Совершенствовать и развивать свой интеллектуальный и общекультурный уровень. Проявлять инициативу, в том числе в ситуациях риска, брать на себя всю полноту ответственности.	Требования ФГОС 3+ (ОК-1,6; ОПК-2; ПК-1,2), критерий 5 АИОР (п. 2.4, п. 2.5), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.
Р11	Демонстрировать способность к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности, способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности, способность к педагогической деятельности.	Требования ФГОС 3+ (ОК-3,4,7; ОПК-3; ПК-7), критерий 5 АИОР (п. 2.6), соответствующий международным стандартам EUR-ACE и FEANI. Запросы студентов, отечественных и зарубежных работодателей.

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

Институт Кибернетики
 Направление подготовки 09.04.01 Информатика и вычислительная техника
 Кафедра Информационных систем и технологий

УТВЕРЖДАЮ:
 Зав. кафедрой

 (Подпись) (Дата)

Мальчуков А.Н.
 (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

Студенту:

Группа	ФИО
8ВМ5А	Чернета Дмитрию Сергеевичу

Тема работы:

Статистические методы оценки качества обучения студентов ТПУ	
Утверждена приказом директора (дата, номер)	От 20.02.2017 №898/с

Срок сдачи студентом выполненной работы:	13.06.2017
--	------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	Выполнения оценки качества обучения студентов ТПУ с использованием статистических методов.
Перечень подлежащих исследованию, проектированию и разработке вопросов	Изучение кластерного, дисперсионного и регрессионного анализа. Разбор критериев Манна – Уитни, Краскела-Уоллиса, Колмогорова-Смирнова, Шапиро-Уилка, Лиллиефорса и обобщенного дискриминантного анализа, проведение первичного анализа исходных данных, проверки первоначальных данных на нормальность распределения, кластерного анализа, проверки равнозначности сложности тем, сравнения данных экзамена за 2015 и 2016 года, проведения регрессионного анализа расчет ресурсоэффективности и ресурсосбережения,

	анализ вредных производственных факторов.
Перечень графического материала	Скриншоты интерфейса программы, диаграммы, графики, гистограммы.
Консультанты по разделам выпускной квалификационной работы	
Раздел	Консультант
Социальная ответственность	Извеков Владимир Николаевич
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Конотопский Владимир Юрьевич
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
Введение, описание предметной области, методы статистического анализа, реферат.	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	10.09.2016
---	------------

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кацман Юлий Янович	Кандидат технических наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ5А	Чернета Дмитрий Сергеевич		

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

Институт Кибернетики
 Направление подготовки 09.04.01 Информатика и вычислительная техника
 Уровень образования Магистратура
 Кафедра Информационных систем и технологий
 Период выполнения осенний / весенний семестр 2016/2017 учебного года

Форма представления работы:

Магистерская диссертация

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
 выполнения выпускной квалификационной работы**

Срок сдачи студентом выполненной работы:	13.06.2017
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
10.11.2016	Первичный анализ данных и их проверка на нормальность распределения	10
15.12.2016	Проверка равнозначности сложности тем	15
06.03.2017	Проведение кластерного анализа	25
14.04.2017	Проведение регрессионного анализа	15
10.05.2017	Сравнение данных экзамена за 2015 и 2016 года	15
1.06.2017	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
5.06.2017	Социальная ответственность	10

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент кафедры ВТ	Кацман Юлий Янович	к.т.н., доцент		

СОГЛАСОВАНО:

Зав. кафедрой	ФИО	Ученая степень, звание	Подпись	Дата
Информационных систем и технологий	Мальчуков Андрей Николаевич	к.т.н., доцент		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8ВМ5А	Чернета Дмитрию Сергеевичу

Институт	Кибернетики	Кафедра	Информационных систем и технологий
Уровень образования	Магистрант	Направление/специальность	09.04.01 Информатика и вычислительная техника

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

<i>1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	
<i>2. Нормы и нормативы расходования ресурсов</i>	
<i>3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<i>1. Оценка коммерческого и инновационного потенциала НТИ</i>	
<i>2. Разработка устава научно-технического проекта</i>	
<i>3. Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок</i>	
<i>4. Определение ресурсной, финансовой, экономической эффективности</i>	

Перечень графического материала (с точным указанием обязательных чертежей):

<i>1. График проведения и бюджет НТИ</i>
<i>2. Оценка ресурсной, финансовой и экономической эффективности НТИ</i>
<i>3. Потенциальные риски</i>

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. менеджмента	Конотопский Владимир Юрьевич	к.э.н., доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ5А	Чернета Дмитрий Сергеевич		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8ВМ5А	Чернета Дмитрию Сергеевичу

Институт	ТПУ ИК	Кафедра	Информационных систем и технологий
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Статистические методы оценки качества обучения студентов ТПУ.
--	---

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<p>1. Производственная безопасность</p> <p>1.1. Анализ выявленных вредных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности:</p> <ul style="list-style-type: none"> – 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). <p>1.2. Анализ выявленных опасных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности:</p> <ul style="list-style-type: none"> – 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). 	<p>Работы ведутся в помещении с уровнем освещения, который соответствует норме в 200-300 люкс (люминесцентные лампы с суммарным уровнем освещенности 215.4 люкс в соответствии со СНиП 23-05-95). Уровень шума в пределах нормы и составляет 37 Дб. Температура воздуха на рабочем месте составляет +24⁰С, относительная влажность 54%, скорость ветра 0.1 мм/с, все параметры соответствуют оптимальным значениям. Аппаратуры для измерения электромагнитного излучение не имеется, потому уровень электромагнитного поля измерен не был.</p> <p>Опасность поражения электрическим током. С целью защиты от поражения электрическим током, возникающим между корпусом приборов и инструментом при пробое сетевого напряжения на корпус, корпуса приборов и инструментов должны быть заземлены, согласно СаНПиН 2.2.2/2.4.1340-03.</p> <p>Опасность возникновения пожара. Для предотвращения пожара помещение с ПЭВМ оборудовано первичными средствами пожаротушения: углекислотным огнетушителем типа ОУ-2 или ОУ-5. Также помещение оснащено</p>
---	---

	пожарной сигнализацией. Рекомендуется оборудовать помещение автоматической установкой объемного газового пожаротушения, например системой азотного пожаротушения «Гарсис».
2. Экологическая безопасность: <ul style="list-style-type: none"> – защита селитебной зоны – анализ воздействия объекта на атмосферу (выбросы); – 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. 	В результате анализа воздействия объекта на литосферу, люминесцентные лампочки являются основным источником загрязнения литосферы.
3. Безопасность в чрезвычайных ситуациях: <ul style="list-style-type: none"> – 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. 	Наиболее типичная чрезвычайная ситуация – пожар. Для предотвращения пожара помещение с ПЭВМ оборудовано первичными средствами пожаротушения: углекислотным огнетушителем типа ОУ-2 или ОУ-5, согласно СНиП 21 – 01 – 9. В случае возникновения ЧС как пожар, необходимо предпринять меры по эвакуации из помещения в соответствии с планом эвакуации, которые расположены на каждом этаже здания, согласно СНиП 21 – 01 – 9.
4. Правовые и организационные вопросы обеспечения безопасности: <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. 	Для предотвращения профессиональных заболеваний необходимо соблюдать правильную рабочую позу при работе за компьютером, иметь максимально регулируемая мебель, которую можно адаптировать к конкретной работе с определенной техникой и периферией, под конкретные физические параметры человека, правильно расположить монитор и клавиатуру.

Дата выдачи задания для раздела по линейному графику	
---	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Извеков В.Н.	к.т.н., доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ5А	Чернета Дмитрий Сергеевич		

РЕФЕРАТ

Выпускная квалификационная работа 143 с., 136 рис., 23 табл., 38 источников.

Ключевые слова: кластерный анализ, критерий Манна – Уитни, критерий Краскела-Уоллиса, регрессионный анализ, нормальность распределения, однофакторный дисперсионный анализ.

Объектом исследования является оценка равнозначности контрольно-измерительных материалов по дискретной математике.

Целью работы является изучение, анализ и исследование результатов экзамена по дискретной математике, используя кластерный, дисперсионный и регрессионный анализ.

Было произведено изучение литературы, исследование разных статистических методов.

В результате исследования были выявлены вопросы трёх уровней сложности, проверены сложности тем, произведено сравнение результатов экзамена за 2015 и 2016 года и построена регрессионная модель.

Работа состоит из следующих разделов.

Первый раздел включает в себя предметную область, роль работы в этой предметной области, и содержит первоначальные материалы для анализа.

Во втором разделе описываются методы статистического анализа, которые были использованы в работе.

В третьем разделе проводится первичный анализ исходных данных.

В четвертом разделе выполняется проверка первоначальных данных на нормальность распределения с целью выбора пригодности различных приёмов статистического анализа.

Пятый раздел содержит кластерный анализ и его результаты.

Шестой раздел содержит информацию о проверке равнозначности сложности тем.

Седьмой раздел содержит в себе сравнение данных экзамена за 2015 и 2016 год.

Восьмой раздел включает в себя регрессионный анализ.

В девятом разделе говорится о финансовом менеджменте, ресурсоэффективности и ресурсосбережении.

В десятом разделе говорится о социальной ответственности.

Область применения: использование итогов исследования для усовершенствования модели оценивания знаний студентов ТПУ.

В дальнейшем предполагается изучение данных с использованием разных нестандартных моделей.

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ, СОКРАЩЕНИЯ

Дисперсия – математическое ожидание квадрата отклонения случайной величины от ее математического ожидания.

Корреляционная матрица – матрица коэффициентов корреляции нескольких случайных величин.

Нормальное распределение – распределение вероятностей, которое описывается плотностью $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ (для непрерывной случайной величины).

Ранг – положение в упорядоченном ряду значений. Значения в некотором измерении сообщают только положение этого значения относительно всех других, однако ничего не говорит о расстоянии между значениями.

Стандартное отклонение – показатель рассеивания значений случайной величины относительно её математического ожидания.

Уровень значимости – достаточно малая вероятность, при которой событие можно считать практически невозможным.

r – коэффициент корреляции.

p -level – уровень значимости.

Multiple R – коэффициент множественной корреляции.

R^2_{adj} – (скорректированный) коэффициент детерминации.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	17
1 ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ	19
2 МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА	20
3 ПОДГОТОВКА И ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ	33
4 ПРОВЕРКА ДАННЫХ НА НОРМАЛЬНОСТЬ РАСПРЕДЕЛЕНИЯ	38
5 КЛАСТЕРНЫЙ АНАЛИЗ	48
5.1 Кластерный анализ	48
5.2 Ранговый критерий Краскела-Уоллиса	51
5.3 Критерий Манна – Уитни	51
5.4 Однофакторный дисперсионный анализ	52
6 ПРОВЕРКА РАВНОЗНАЧНОСТИ СЛОЖНОСТИ ТЕМ	54
6.1 Анализ данных	54
6.2 Ранговый критерий Краскела-Уоллиса	56
6.3 Критерий Манна – Уитни	57
6.4 Группировка тем	60
6.5 Однофакторный дисперсионный анализ	62
7 СРАВНЕНИЕ ДАННЫХ ЭКЗАМЕНА ЗА 2015 И 2016 ГОД	65
7.1 Обработка новых данных	65
7.2 Сравнение данных	67
7.2.1 Критерий Манна – Уитни	68
7.2.2 Критерий Краскела-Уоллиса	72
8 РЕГРЕССИОННЫЙ АНАЛИЗ	77
8.1 Корреляционный анализ	77
8.2 Линейный регрессионный анализ	79
8.3 Экспоненциальный регрессионный анализ	81
8.4 Кусочно-линейный регрессионный анализ с точкой останова	83
8.5 Обобщенный дискриминантный анализ	84
9 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ	88

9.1 Организация и планирование работ	88
9.1.1 Продолжительность этапов работ	89
9.1.2 Расчет накопления готовности проекта	93
9.2 Расчет сметы затрат на выполнение проекта	94
9.2.1 Расчет затрат на материалы	94
9.2.2 Расчет заработной платы	94
9.2.3 Расчет затрат на социальный налог	95
9.2.4 Расчет затрат на электроэнергию	95
9.2.5 Расчет амортизационных расходов	96
9.2.6 Расчет расходов, учитываемых непосредственно на основе платежных (расчетных) документов (кроме суточных)	97
9.2.7 Расчет прочих расходов	98
9.2.8 Цена разработки ВКР	98
9.2.9 Прибыль	98
9.2.10 Расчет НДС	98
9.2.11 Цена разработки ВКР	98
9.3 Оценка экономической эффективности проекта	99
9.4 Оценка научно-технического уровня ВКР	99
10. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ	101
10.1 Микроклимат помещения	102
10.2 Освещение.....	106
10.3 Производственный шум	109
10.4 Электромагнитное излучение	112
10.5 Электрическая безопасность	115
10.6 Экологическая безопасность.	115
10.7 Безопасность в чрезвычайных ситуациях.	116
10.8 Правовые и организационные вопросы обеспечения безопасности.	118
ЗАКЛЮЧЕНИЕ	121
СПИСОК ПУБЛИКАЦИЙ СТУДЕНТА	122
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	123

Приложение А	127
ПРИЛОЖЕНИЕ В CD-ДИСК	143

ВВЕДЕНИЕ

Оценка равнозначности контрольно-измерительных материалов, используемых для оценивания знаний студентов, является актуальной задачей и в новых дисциплинах, и в тех, которые преподаются уже много лет. Контрольно-измерительные материалы, как правило, представлены в нескольких вариантах, из-за чего возникает проблема их равнозначной сложности, что может усложнить оценку знаний студентов и ее объективность.

Обычно при анализе свойств проверяющих заданий много внимания тратится на проверку параллельности вариантов заданий. Тем временем если для использования современной теории тестов – Item Response Theory (IRT) при оценивании скрытых факторов необходимо представить для каждого теста выборку минимум от 200 до 1000 опытов, то классическая статистическая теория даёт возможность получить анализ параметров, при значительно меньшем количестве опытов.

Объектом исследования является оценка равнозначности контрольно-измерительных материалов по дискретной математике. Предметом исследования является кластерный, дисперсионный и регрессионный анализ экзаменационных заданий по дискретной математике. Практическая значимость заключается в нахождении возможных вопросов и тем различной сложности и дальнейшего распределения на экзамене между студентами вопросов одинаковой сложности, чтобы оценка за экзамен зависела от знания студентом материала, а не от сложности вопроса.

Целью работы является изучение, анализ и исследование результатов экзамена по дискретной математике, используя кластерный, дисперсионный и регрессионный анализ.

Были поставлены следующие задачи:

- произвести подготовку данных для дальнейшего анализа;
- сделать проверку данных на нормальность распределения;
- разделить вопросы по уровню сложности;
- проверить равнозначность сложности тем;

- сравнить полученные баллы за 2015 и 2016 года;
- попытаться предсказать результаты экзамена по баллам за один вопрос.

1 ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ

Изучение закономерностей, по которым происходят массовые случайные явления, базируется на анализе итогов наблюдений при помощи приёмов статистического анализа.

Задача номер один математической статистики – это определить методы собирания и объединения статистических сведений, которые были получены при помощи наблюдений, либо же при помощи специально организованных экспериментов.

Задача номер два математической статистики – исходя из целей исследования, выработать способы анализа статистических данных:

- Анализ неопределённой вероятности события; Анализ неопределённой функции распределения; Анализ параметров известного распределения; Анализ корреляции случайной величины от других случайных величин и т.д.;
- Проверка гипотез статистики о характере неизвестного распределения, или о размере распределения параметров, характер которого известен.

В наше время математическая статистика создаёт методы установления количества необходимых испытаний ещё перед началом исследования, во время исследования, и выполняет другие задачи. Нынешнюю математическую статистику характеризуют как дисциплину об утверждении решений в ситуации неопределенности.

Цель математической статистики заключается в формировании способов сбора и обработки данных статистики с целью извлечения научных и практических заключений.

Для исследования были даны оценки за экзамен по дискретной математике, включающие в себя данные за 2015 и 2016 года.

2 МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА

Есть разные методы статистического анализа, созданные для определения конкретных зависимостей и разрешения разных задач математической статистики.

В текущей работе были использованы следующие методы статистического анализа:

- кластерный анализ;
- дисперсионный анализ;
- регрессионный анализ.

Так же была проведена проверка данных на нормальность распределения, потому что большое число методов статистики вытекает из предположения нормальности распределения исследуемых данных.

Кроме того в работе использовались:

- Критерий Манна – Уитни;
- Ранговый критерий Краскела-Уоллиса;
- Обобщенный дискриминантный анализ;
- Критерий Колмогорова-Смирнова;
- Критерий Шапиро-Уилка;
- Критерий Лиллиефорса.

1.1 Кластерный анализ

Кластерный анализ является методом классификационного анализа, его главная цель – это разделение большого количества исследуемых объектов на однородные группы, они же кластеры.

Задача кластерного анализа состоит в том, чтобы в соответствии с признаками объектов разделить их на m кластеров так, чтобы каждый объект относился только одному кластеру. Вместе с тем объекты, принадлежащие одной группе, должны быть схожими, а объекты, принадлежащие разным группам, – несхожими.

При представлении объектов кластеризации в виде точек в n-мерном пространстве признаков, похожесть объектов можно обозначить через расстояние между точками, и интуитивно понятно, что чем более схожи объекты, тем меньше расстояние между ними. Существуют разные способы вычисления расстояния между точками.

Квадрат евклидова расстояния. Представляет собой расстояние в многомерном пространстве и рассчитывается так:

$$\text{Расстояние}(x, y) = \sum_i (x_i - y_i)^2 \quad (1)$$

Квадрат евклидова расстояния рассчитывается из исходных, а не из стандартизованных данных. Этот метод его расчета имеет конкретные преимущества (к примеру, расстояние меж двух объектов не меняется при добавлении в модель новых данных, которые могут быть выбросом). Однако отличия между осями могут значительно воздействовать на расстояния, по координатам которых они рассчитываются. Например, если перевести одну из осей из сантиметров в миллиметры, (увеличив её значения в 10 раз), то финальный квадрат евклидова расстояния, рассчитываемый по координатам, значительно поменяется, и, следовательно, финальные значения кластерного анализа могут значительно измениться.

Манхэттенское расстояние представляет собой среднее координатных разностей. В основном это расстояние показывает такие же результаты, что и Евклидово расстояние. Стоит заметить, что для этой меры воздействие выбросов менее значимо (в виду того, что не нужно их возводить в квадрат). Формула для вычисления Манхэттенского расстояния:

$$\text{Расстояние}(x, y) = \sum_i |x_i - y_i| \quad (2)$$

Расстояние Чебышева будет особенно нужно, когда нужно идентифицировать два объекта как «разные», если эти объекты различаются по одной координате. Формула для вычисления расстояние Чебышева:

$$\text{Расстояние}(x, y) = \text{Максимум}|x_i - y_i|$$

Степенное расстояние позволяет постепенно увеличивать или уменьшать вес размерности, в которой подобные объекты значительно различаются. Формула степенного расстояния:

$$\text{Расстояние}(x, y) = \left(\sum_i |x_i - y_i|^p \right)^{\frac{1}{r}} \quad (3)$$

где r и p — пользовательские параметры. Переменная p отвечает за постепенное определение веса разностей по разным координатам, переменная r отвечает за прогрессивное определение веса значительных расстояний между объектами. В расстоянии Евклида переменные r и p , равны двум.

Процент несогласия применяется тогда, когда объекты представляют собой категориальные данные. Формула вычисления расстояния: [1]

$$\text{Расстояние}(x, y) = \frac{(\text{Количество } x_i \neq y_i)}{i} \quad (4)$$

Правила объединения или связи

Сначала, когда любой объект является отдельным кластером, выбранная мера определяют расстояния между объектами. Между тем, когда объединяются вместе множество объектов, появляется вопрос, как можно найти расстояния между кластерами? Иными словами, нужен метод группировки для двух кластеров. Есть ряд методов группировки кластеров.

В методе **одионой связи**, которая так же называется **методом ближайшего соседа**, расстояние между двумя кластерами определяется расстоянием между двумя самыми близкорасположенными объектами в разных кластерах. Данное правило, объединяет объекты вместе для создания кластеров, и финальные кластеры обладают склонностью состоять из протяжённых "цепочек".

В методе **полной связи**, которая так же называется как **метод наиболее удаленных соседей**, расстояние между кластерами задаётся максимальным расстоянием между какими угодно двумя объектами из разных кластеров. Данный способ, как правило, действует отлично, когда объекты берут начало на самом деле из действительно разных "роц". В случае же когда кластеры

обладают продолговатой формой или их настоящий вид представляет собой "цепочным", то данным способ использовать нельзя.

В методе **невзвешенного попарного среднего** расстояние между двумя разными кластерами рассчитывается как среднее расстояние между всеми парами объектов в них. Способ результативен, когда объекты на самом деле создают разные "рощи", но он функционирует так же качественно и в случаях длинных ("цепочного" типа) кластеров.

Метод **взвешенного попарного среднего** похож на метод невзвешенного попарного среднего, кроме того, что при расчетах величина подобных кластеров (т.е. количество объектов, которое они содержат) применяется в роли весового коэффициента. Следовательно, данный метод следует применять, если есть предположения, что кластеры имеют разные размеры.

Во **невзвешенном центроидном методе** расстояние между двумя кластерами рассчитывается как расстояние между их центрами тяжести.

Взвешенный центроидный метод совпадает с предыдущим, кроме того, что при расчёте применяются веса для расчёта разницы между объёмом кластеров (т.е. числом элементов в них). Соответственно, если есть значимые различия в объёмах кластеров, то лучше использовать данный метод, чем предыдущий.

Метод Варда рознится от иных способов, так как он для расчёта расстояний между кластерами применяет методы дисперсионного анализа. Способ сводит к минимуму сумму квадратов (SS) для любых двух кластеров, которые могут быть созданы на любом шаге. По большому счёту метод является крайне действенным, но он тяготеет к созданию кластеров малого размера. [2]

1.2 Дисперсионный анализ

Метод дисперсионного анализа заключается в нахождении доли систематической (межгрупповой) дисперсии к случайной (внутригрупповой) дисперсии в замеренных данных. В виде индикатора изменчивости

применяется сумма квадратов отклонения значений параметра от среднего: SS. Общая сумма квадратов SSTotal расчлняется на межгрупповую сумму квадратов SSBG и внутригрупповую сумму квадратов SSWG:

$$SSTotal = SSBG + SSWG$$

Если истинна гипотеза H_0 , то внутригрупповая и межгрупповая дисперсии являются оценками одной и той же дисперсии и должны быть примерно одинаковы.

$$F = \frac{MS_{BG}^4}{MS_{WG}} \quad (5),$$

$$\text{где } MS_{BG} = \frac{SSBG}{vBG} \text{ и } MS_{WG} = \frac{SSWG}{vWG}$$

Следовательно, значение F должно быть близко к 1, если статистически значимые различия отсутствуют. Значение F определяется уровнем значимости (обычно 0,05 или 0,01). [3]

В программе статистика дисперсионный анализ выглядит следующим образом (рисунок 1).

Analysis of Variance (Voprosi bolshe 4 NEW)								
Marked effects are significant at p < ,05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Ball	5,147886	2	2,573943	22,92651	275	0,083369	30,87406	0,000000

Рисунок 1 – Результаты дисперсионного анализа

где,

SS (Sum of Squares) Effect – сумма квадратов факторов умноженная на k - 1;

df Effect – число степеней свободы фактора;

MS (Mean Square) Effect – средний квадрат фактора;

SS Error – сумма квадратов умноженная на N – k;

df Error – число степеней свободы равная N – k;

MS Error – оценка дисперсии;

F – значение статистики Фишера;

p –вероятность принятия гипотезы H_0 .

1.3 Регрессионный анализ

Регрессия это величина, выражающая зависимость среднего значения некоторой случайной величины от какой-либо другой величины, либо же от нескольких величин.

Она может применяться как для решения прикладных, так и в исследовательских целях, для предсказания некоего результата по некоторым предварительно измеренным характеристикам.

Кроме того, с помощью регрессии можно определять, какие показатели наиболее существенны, и важны для исследования, а какими из переменных можно пренебречь и исключить из анализа.

Регрессию используют для количественного обозначения связей между переменными, и выражают уравнением регрессии. [4]

Главная задача регрессионного анализа заключается в нахождении связи между некоторым признаком Y наблюдаемого объекта и параметрами x_1, x_2, \dots, x_n , которые интерпретируют изменения Y . Признак Y является зависимой переменной (откликом), влияющие параметры x_1, x_2, \dots, x_n являются факторами. Нахождение вида зависимости, нахождение уравнения регрессии и оценивание ее параметров – это задачи регрессионного анализа. [5]

В программе статистика регрессионный анализ выглядит следующим образом (рисунок 2).

1.4 Критерий Манна – Уитни

Критерий Манна-Уитни – непараметрический статистический критерий, употребляемый при сравнении двух независимых выборок по степени любого, измеренного количественно, признака. Способ базируется на распознавании того, довольно ли мала область перекрывающихся показателей между двумя вариативными рядами. Отличия между показателями параметра в выборках тем вернее, чем больше значение критерия. [6]

В программе статистика критерий Манна-Уитни выглядит следующим образом (рисунок 3).

Mann-Whitney U Test (Voprosi bolshe 4 NEW)									
By variable Cluster									
Marked tests are significant at p <.05000									
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2
Ball	12962,00	6541,000	2084,000	-4,56718	0,000005	-4,60304	0,000004	147	50

Рисунок 3 – Результаты теста Манна – Уитни

где,

Rank Sum T_i – сумма рангов выборки T_i ;

U – статистика Манна - Уитни для малых выборок;

Z – нормальная аппроксимация статистики Манна - Уитни для больших выборок;

p - level – вероятность принятия гипотезы H_0 ;

Z adjusted – скорректированная нормальная аппроксимация статистики Манна - Уитни;

p - level – скорректированная вероятность принятия гипотезы H_0 ;

Valid N – объем выборки;

1.5 Ранговый критерий Краскела-Уоллиса

Критерий используется для оценки отличий одновременно между тремя и более выборками. Он даёт возможность выяснить, что признак меняется при смене одной группы на другую, но не показывает на направление изменений. Представляет собой обобщение критерия Манна-Уитни на большее количество выборок.

Все значения группируются и выстраиваются в общем ряду. Потом в каждой выборке рассчитываются суммы рангов. Если отличия оказываются случайными, то высокие и низкие ранги однородно разделятся по выборкам. Когда в какой либо группе будут доминировать высокие ранги, а в оставшейся группе низкие, то это свидетельствует о том, что отличия не случайны, а вызваны действием фактора. [7]

В программе статистика критерий Краскела-Уоллиса выглядит следующим образом (рисунок 4).

Kruskal-Wallis ANOVA by Ranks; Ball (Voprosi bolshe 4 NEW)					
Independent (grouping) variable: Cluster					
Kruskal-Wallis test: H (2, N= 278) =52,37230 p =,0000					
Depend.: Ball	Code	Valid N	Sum of Ranks	Mean Rank	
1	1	147	21091,00	143,4762	
2	2	50	9945,00	198,9000	
3	3	81	7745,00	95,6173	

Рисунок 4 – Результаты теста Краскела - Уоллиса

где,

H – статистика Краскела - Уоллиса;

p – вероятность принятия гипотезы H_0 .

Codes – уникальный код группы;

Valid N – число значений в группе;

Sum of Ranks – сумма рангов;

Mean Rank – среднее значение ранга;

1.6 Обобщенный дискриминантный анализ

При проведении дискриминантного анализа создаётся прогностическая модель для определения принадлежности к группе. Эта модель создаёт набор дискриминантных функций в виде линейного сочетания предикторных переменных, обеспечивающий оптимальное разбиение групп. Данные функции создаются по совокупности наблюдений, для которых известна их принадлежность к группам, и могут далее быть применены к свежей выборке с известными показателями предикторных значений, но неизвестной групповой принадлежностью. [8]

На входе дискриминантного анализа – категориальная переменная, содержащая принадлежность к классам (например, сегментация, или потребление продуктов или брендов), и набор независимых переменных, измеренных по интервальным или дихотомическим шкалам. При этом категориальная переменная может иметь пропуски, т.е. не для всех респондентов будет заполнена.

С помощью статистического критерия алгоритм ищет такую комбинацию линейных уравнений (дискриминантных функций), которая наилучшим образом опишет различия между группами.

Первая дискриминантная функция отделяет первую группу от групп 2, 3, ... N, вторая – вторую группу от групп 3, 4, ... N и т.д. В итоге все N групп получают отделёнными друг от друга дискриминантными функциями. Это даёт возможность определить вероятность принадлежности респондента к тому или иному классу.

Дискриминантный анализ дает возможность понять, по каким именно переменным имеющиеся классы различаются сильнее всего.

Каждому респонденту приписывается вероятность попадания в тот или иной класс; таким образом, при сегментации можно определить основной (по наибольшей вероятности попадания) и дополнительный (второе по величине значение вероятности) сегмент респондента.

Метод позволяет воспроизвести уже имеющуюся сегментацию, распространив ее на новые волны исследования. Кроме того, с его помощью можно оценить устойчивость сегментов.

Наконец, с помощью дискриминантного анализа можно строить различные карты, существенно облегчающие интерпретацию результатов кластерного анализа. [9]

1.7 Критерий Колмогорова-Смирнова

Характеристиками распределения, как правило, служат математическое ожидание и дисперсия. В отдельных случаях взамен дисперсии берут её корень

– стандартное отклонение. Для теста гипотезы применяют следующие критерии: **Колмогорова-Смирнова**, либо же **Шапиро-Уилка**. Между тем при анализе менее 60 наблюдений, как правило, используется критерий Шапиро-Уилка, в противном случае критерий Колмогорова-Смирнова. [10]

Критерий Колмогорова-Смирнова – представляет собой непараметрический критерий согласия, он был создан для того, чтобы проверять, принадлежит ли анализируемая выборка какому-либо закону распределения. Чаще всего этот критерий применяется для анализа выборок на нормальность распределения.

Данный критерий рассчитывается по следующей формуле:

$$D_n = \sup_x |F_n(x) - F(x)|, \quad (6)$$

где \sup S - точная верхняя грань множества S , F_n - функция распределения исследуемой совокупности, $F(x)$ - функция нормального распределения. [11]

Критерий Колмогорова-Смирнова находит место, в котором сумма расхождений двух распределений является максимальной, и проанализировать истинность данного расхождения.

В случае, когда отличия двух распределений значительны, то в один момент разность собранных частот достигнет ключевого показателя, что будет значить, что различия являются статистически достоверными. [12]

1.8 Критерий Шапиро-Уилка

Исследовать сходство выборки нормальному распределению возможно используя критерий Шапиро-Уилка. Данный критерий базируется на Q-Q графике, он отслеживает, как значительно точки на Q-Q графике не совпадают с прямой, анализирует гипотезу H_0 , и совершает это при помощи статистики W , вычисляемой согласно вариационному ряду, рассчитанному по выборке, и некоторых величин a . Эти значения базируются на математических ожиданиях порядковых статистик из стандартного нормального распределения, они взяты из таблиц и для них нет формул.

Тест Шапиро-Уилка применяется для тестирования гипотезы о том, что случайная величина имеет нормальное распределение и служит одним из самых действенных тестов проверки на нормальность. [13]

Критерий вычисляется как отношение квадрата суммы линейной разности выборочных порядковых статистик к обычной оценке дисперсии.

Критерий базируется на упорядоченных данных. Если последовательность из n независимых опытов, размещённую в порядке неубывания, пометить переменными x_1, x_2, \dots, x_n , то можно рассчитать по формуле промежуточную сумму S :

$$S = \sum a_k [x_{(n+1-k)} - x_k], \quad (7)$$

где k - индекс, принимающий значения от 1 до $n/2$ или от 1 до $(n-1)/2$ при четном и нечетном n соответственно;

a_k - коэффициент, обладающий особыми значениями для размера выборки n .

Тогда статистика величины W выглядит следующим образом:

$$W = S^2 / (nm_2),$$

$$\text{где } nm_2 = \sum (x_i - \bar{x})^2;$$

n - объем выборки;

m_2 - выборочный центральный момент второго порядка.

Если величины каких-либо опытов одинаковы, то упорядоченная последовательность нумеруется с копированием одинаковых опытов столько раз, сколько раз они появляются в исходной группе. [14]

1.9 Критерий Лиллиефорса

Критерий Лиллиефорса — представляет собой статистический критерий, который является вариацией критерия Колмогорова – Смирнова. Применяется с целью теста нулевой гипотезы о том, что данные имеют нормальный закон распределения для того варианта, когда математическое ожидание и дисперсия нормального распределения заведомо неизвестны.

Как и в случае с критерием Колмогорова–Смирнова, вычисляется максимальная разница среди выборочной и теоретической интегральными функциями распределения.

Выполняется проверка, окажется ли статистически важным рассматриваемое смещение выборочной функции распределения от теоретической. При возникновении положительного решения, нулевая гипотеза отклоняется.

Главная причина погрешности критерия Лиллиефорса заключается в том, что параметры теоретического распределения считаются по тем же значениям, которые проверяются на соответствие распределению. То есть, наибольшее отклонение будет меньше, чем, если параметры распределения вычисляются независимо. Следовательно, распределение вероятности гипотезы о верности нулевой гипотезы, становится сдвинуто в сторону меньших значений в сравнении с распределением Колмогорова. [15]

$$\text{Статистика критерия } D^* = \max|\hat{F}(x) - G(x)|, \quad (8)$$

где

$\hat{F}(x)$ – эмпирическая функция распределения

$G(x)$ - нормальное распределение, построенное по выборочному среднему и выборочной дисперсии. [16]

3 ПОДГОТОВКА И ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ

В качестве исходных данных использовалась таблица, содержащая в себе номера вариантов, фамилии и количество баллов за каждый из шести вопросов из экзамена по дискретной математике (таблица 1).

Таблица 1 – Исходная таблица

Вариант	Фамилия	Вопрос 1	Вопрос 2	Вопрос 3	Вопрос 4	Вопрос 5	Вопрос 6	Сумма
2	Комаров Вячеслав	5	3	3	5	4	3	23
12	Кузиванов Николай	4	6	6	5	2	6	29
5	Авдеев Николай	4	4	5	7	5	5	30
11	Журавлев Сергей	3	2	2	0	3	2	12
27	Комаров Игорь	4	5	4	4	7	7	31
10	Кузнецова Анастасия	6	5	6	7	7	7	38
20	Виноградов Кирилл	2	3	2	2	2	2	13
18	Голуб Алексей	4	2	2	7	1	2	18
13	Иовлев Юлий	6	2	6	4	7	3	28
24	Козлова Александра	4	3	6	7	7	7	34
29	Горюнова Марина	6	6	3	7	3	6	31
16	Квасникова Ирина	6	6	4	7	4	7	34
28	Белков Сергей	5	5	3	7	5	0	25
9	Шумилин Олег	4	3	3	0	7	7	24
21	Чиликин Иван	5	6	7	7	7	7	39
10	Голуб Алексей	3	5	3	3	7	7	28
17	Соколова	3	3	5	0	7	2	20
26	Журавлев Сергей	3	0	2	3	3	2	13
24	Черепанов	4	2	7	2	6	0	21
19	Удалов	2	1	2	0	1	7	13
20	Курилова	4	1	7	6	5	4	27
20	Фирсанков	4	4	3	7	7	1	26
16	Первушина	5	3	3	1	4	2	18
13	Малюгин	0	1	1	2	4	2	10
27	Новицкий	1	1	0	0	4	2	8
8	Черемнов	3	5	3	5	0	2	18
9	Мозгалева	3	0	1	0	7	6	17
10	Реннит	2	0	2	0	7	0	11
21	Гриценко	2	2	2	7	7	2	22
25	Гусева	3	2	4	1	7	0	17
26	Коломейцева	3	6	5	0	3	1	18

5	Ткачев	6	3	5	6	2	1	23
29	Леухин	1	0	3	0	7	3	14
12	Лысенко	4	5	7	6	2	0	24
11	Башлыков	4	0	2	1	1	3	11
1	Рот	3	4	3	3	7	7	27
22	Каун	4	2	6	3	7	0	22
7	Самушкина	3	5	4	0	7	1	20
23	Исунц	4	3	1	2	6	3	19
3	Старшинов	4	5	5	5	4	5	28
4	Чайбар	0	4	2	0	6	0	12
24	Виноградов	1	0	5	6	3	2	17
15	Хомеченко	3	3	6	3	7	1	23
14	Семенов	1	0	3	1	3	7	15
17	Черепов	6	2	5	4	4	1	22
18	Никифоров	5	2	0	0	5	3	15
2	Катышева	6	4	6	7	7	7	37
19	Шкабара	3	0	7	3	7	7	27
14	Русанович	3	3	7	3	1	7	24
17	Толстых	6	3	7	4	7	1	28
25	Феклин	3	0	1	1	0	3	8
3	Колочев	3	3	3	4	0	0	13
23	Нагиев	3	3	3	4	3	7	23
26	Ягунов	5	3	6	4	3	3	24
22	Перминов	6	4	7	1	0	2	20
15	Гусева	3	2	2	3	7	2	19
24	Никифоров	2	0	7	4	3	3	19
22	Пономарев	4	4	6	1	5	3	23
1	Коломейцева	1	2	4	5	3	0	15
27	Сидоров	6	3	6	0	7	2	24
17	Дресвянских	6	6	7	5	7	6	37
28	Феклин	3	3	3	3	4	0	16
20	Колочев	3		7	3	3	5	
25	Костюченко	6	5	7	7	7	5	37
7	Лысенко	2	4	3	0	2	5	16
18	Чайбяр	3	0	1	0	3	4	11
2	Агеева	3	4	4	5	4	4	24
23	Черепов	4	4	3	0	4	0	15
3	Парфенов	5	2	3	4	4	5	23
29	Перминов				5	7	7	
10	Голуб	3	5	3	3	7	7	28

Далее, для дальнейшей работы, баллы за вопросы в исходной таблице были переведены в проценты. Потом, на основе исходной таблице была составлена таблица, отображающая распределение баллов по вариантам. Затем была составлена таблица вопросов, показывающая, какое количество баллов получили студенты за каждый вопрос. Одинаковые вопросы были объединены. Далее из таблицы вопросов, для дальнейшей работы, были убраны вопросы, на

которые имелось менее 4 ответов. Была получена таблица 2, которую уже можно статистически анализировать.

Таблица 2 – Таблица баллов за каждый вопрос, с количеством ответов не менее 4

№ вопроса	Баллы за вопрос																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	0,5	1	0,8 33 33 3	0,3 33 33 3	0,6 66 66 7	0,6 66 66 7	0,3 33 33 3	0,6 66 66 7											
3	0,2 85 71 4	1	0,2 85 71 4	0,2 85 71 4	0,4 28 57 1	0,5 71 42 9	0,1 42 85 7	0,8 57 14 3	0	1	0,4 28 57 1	0,5 71 42 9	0,5 71 42 9	0,4 28 57 1	0,1 42 85 7	0,4 28 57 1			
4	0,2 85 71 4	0,5 71 42 9	0,2 85 71 4	0,4 28 57 1	0	0,4 28 57 1	0,4 28 57 1	1	0,1 42 85 7	0,4 28 57 1	0,7 14 28 6	1	0						
7	0,8 33 33 3	0,5	1	1	1	0,5													
8	0,4 28 57 1	0,7 14 28 6	0,5 71 42 9	0,2 85 71 4	0,7 14 28 6	0,8 57 14 3	0,5 71 42 9												
9	0,7 14 28 6	0,7 14 28 6	0,7 14 28 6	1	0,5 71 42 9	0,7 14 28 6	0,5 71 42 9												
12	0,5	0,6 66 66 7	0,5	0,5	1	0,8 33 33 3													
13	0,3 33 33 3	0,5	0,1 66 66 7	0	0,8 33 33 3	0,6 66 66 7	0,5	0,3 33 33 3											
14	0,8 57 14 3	0,5 71 42 9	0,2 85 71 4	0,4 28 57 1	0,7 14 28 6	1	0,4 28 57 1	0	0,4 28 57 1	1	0,7 14 28 6	0,7 14 28 6	1	0,4 28 57 1	0,8 57 14 3	1	1	0,4 28 57 1	0
17	0,5	0,6 66 66 7	0,6 66 66 7	0,6 66 66 7	0	0,1 66 66 7	0,3 33 33 3	1											
18	0,8 57 14 3	0,8 57 14 3	0,5 71 42 9	0,2 85 71 4	1	0,4 28 57 1	0,1 42 85 7	0,7 14 28 6	0,8 57 14 3	0,2 85 71 4	0,7 14 28 6	0,8 57 14 3	1	1	0,8 57 14 3				

22	0,6 66 66 7	1	0,8 33 33 3	1															
23	0,6 66 66 7	0,3 33 33 3	1	0,8 33 33 3	0,5	0,5	0	0,5											
24	1	0,5 71 42 9	1	0,4 28 57 1	0	0	0,8 57 14 3	0,4 28 57 1	0,1 42 85 7	0,1 42 85 7	0	0,4 28 57 1							
38	0,6 66 66 7	1	1	0,6 66 66 7	0	0,5	0,1 66 66 7	0,8 33 33 3	0,5	0									
39	0,4 28 57 1	0	0,1 42 85 7	1	0,5														
40	0,7 14 28 6	0,8 57 14 3	0,4 28 57 1	0	0,4 28 57 1	0	0,2 85 71 4	0,8 57 14 3	0,4 28 57 1	0,4 28 57 1									
43	1	0,5	0,3 33 33 3	0,3 33 33 3	0,5	0,5													
44	0,8 33 33 3	0,8 33 33 3	0	0,8 33 33 3															
45	1	1	1	1															
46	1	0,2 85 71 4	1	0,5 71 42 9	0,1 42 85 7	0	0,7 14 28 6	1											
47	0	0	0,1 42 85 7	0,1 42 85 7	0,4 28 57 1	0,4 28 57 1													
50	0,6 66 66 7	0,3 33 33 3	0,6 66 66 7	0,6 66 66 7	0,6 66 66 7	0,5													
52	0,2 85 71 4	0,4 28 57 1	0,4 28 57 1	0,2 85 71 4	0,4 28 57 1														
53	0,8 57 14 3	1	0	0	1														
55	0,8 33 33 3	0,3 33 33 3	0,6 66 66 7	0,1 66 66 7	0,5	0,5	0,6 66 66 7												
56	0,5	0,3 33 33	0,3 33 33	0	0,3 33 33	0,5	0,5	0	0,6 66 66	0,6 66 66	1	0,8 33 33							

4 ПРОВЕРКА ДАННЫХ НА НОРМАЛЬНОСТЬ РАСПРЕДЕЛЕНИЯ

С целью выбора пригодности различных приёмов статистического анализа была проведена проверка данных на нормальность распределения. Для этого, на основе таблицы баллов за каждый вопрос, с количеством ответов не менее 4, для каждого вопроса за 2015 и 2016 года были построены 42 гистограммы (рисунок 5 – 25).

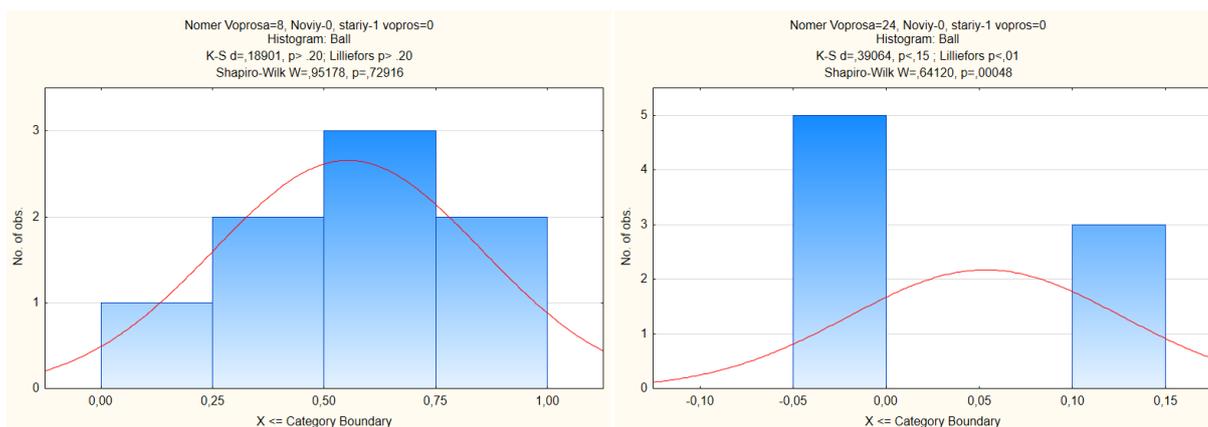


Рисунок 5 – Гистограмма частот для 8 и 24 вопросов за 2016 год

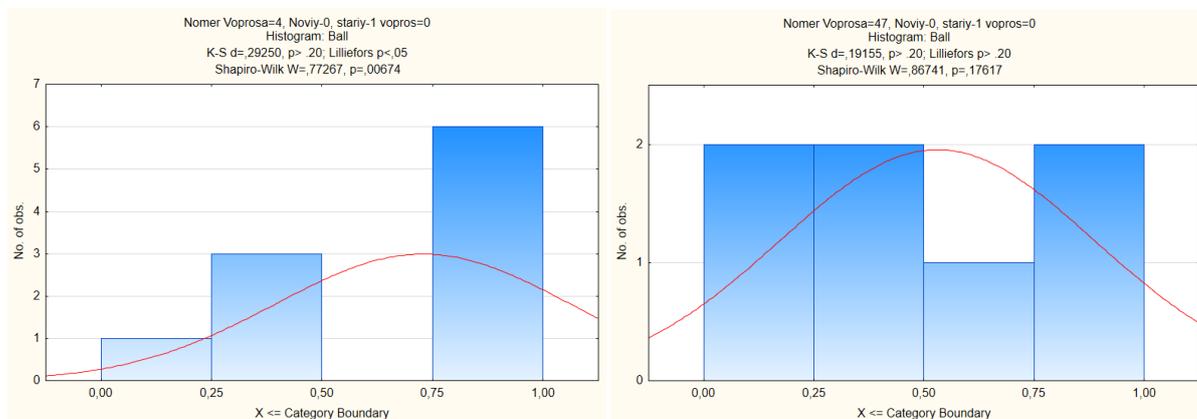


Рисунок 6 – Гистограмма частот для 4 и 47 вопросов за 2016 год

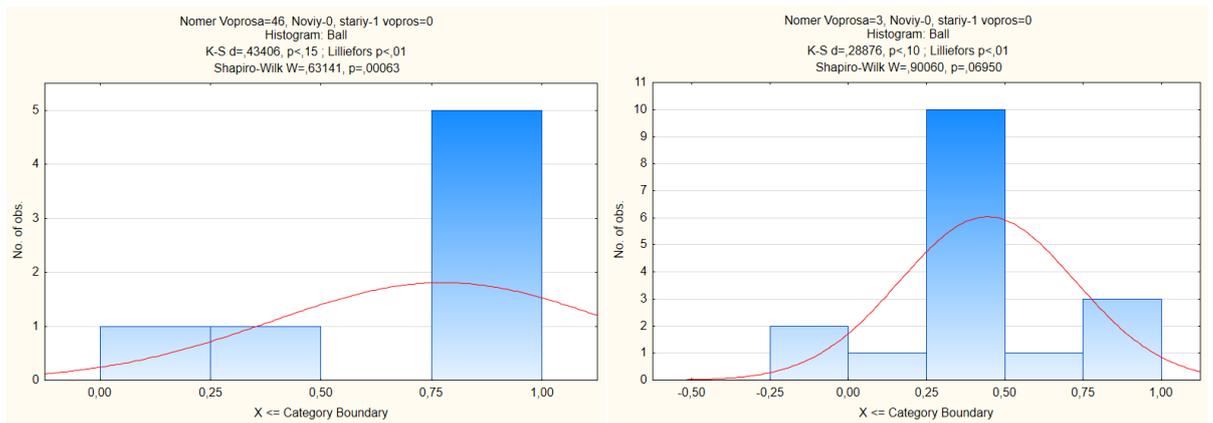


Рисунок 7 – Гистограмма частот для 46 и 3 вопросов за 2016 год

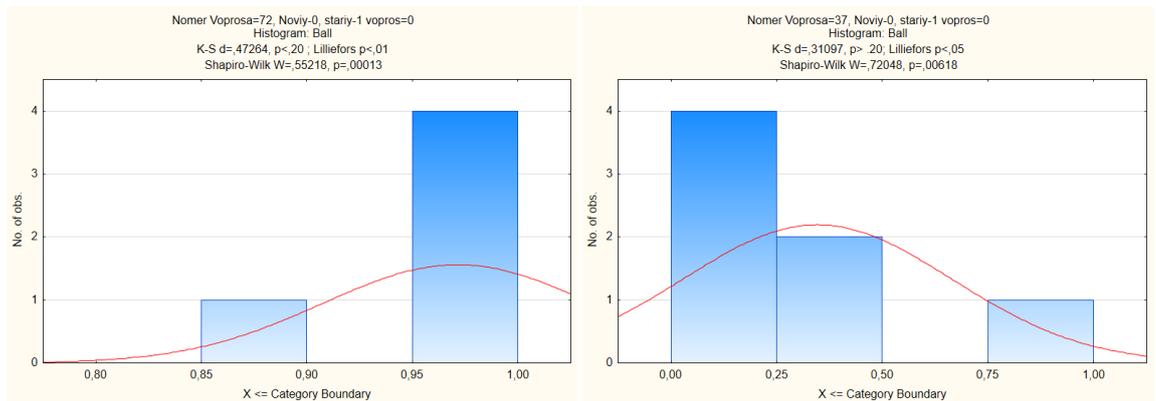


Рисунок 8 – Гистограмма частот для 72 и 37 вопросов за 2016 год

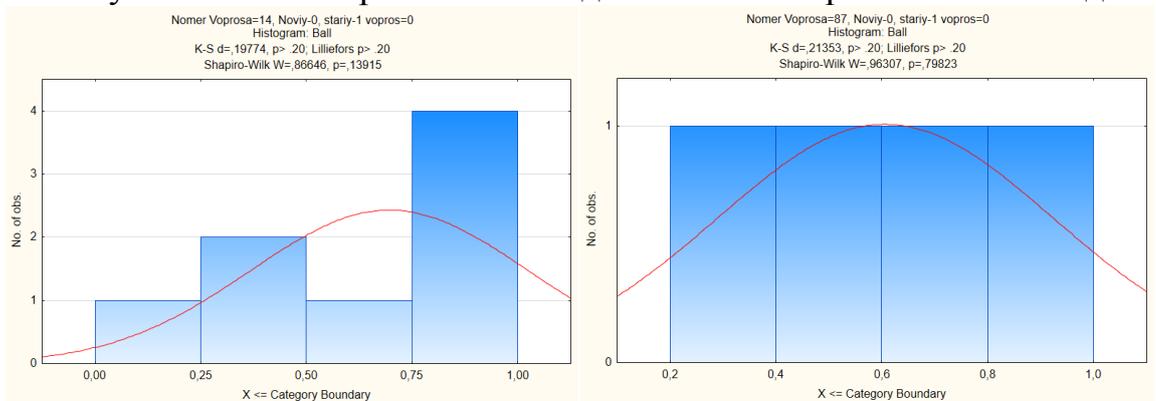


Рисунок 9 – Гистограмма частот для 14 и 87 вопросов за 2016 год

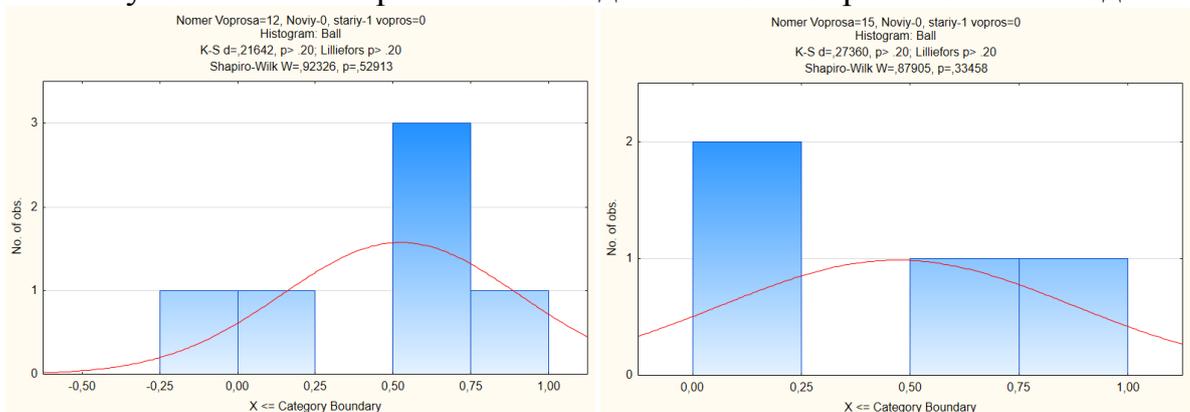
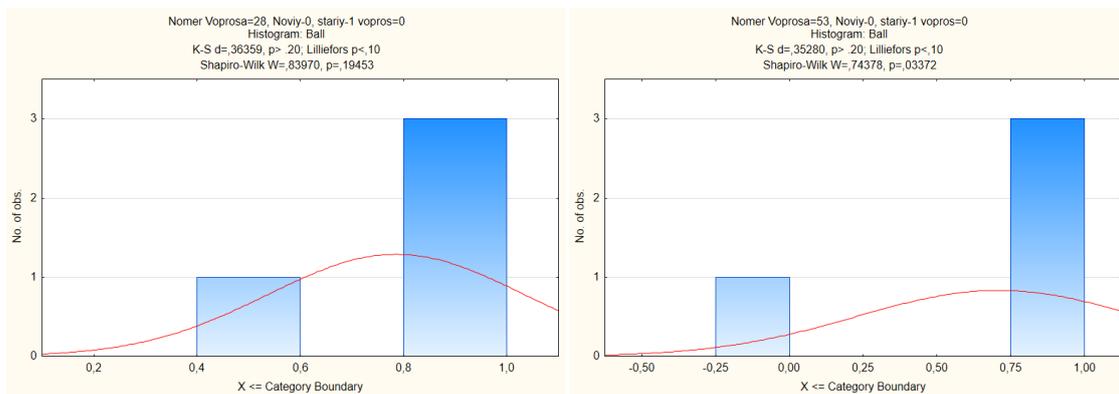
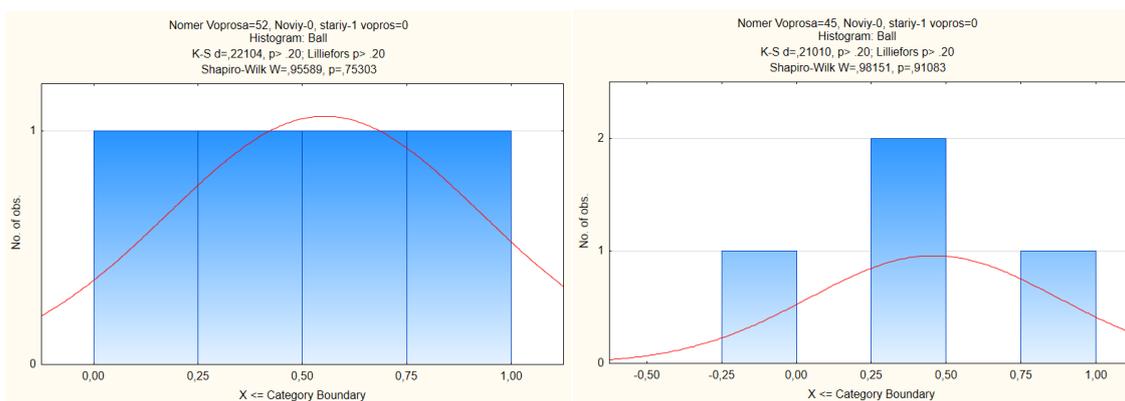


Рисунок 10 – Гистограмма частот для 12 и 15 вопросов за 2016 год



Рисунок

11 – Гистограмма частот для 28 и 53 вопросов за 2016 год



Рисунок

12 – Гистограмма частот для 52 и 45 вопросов за 2016 год

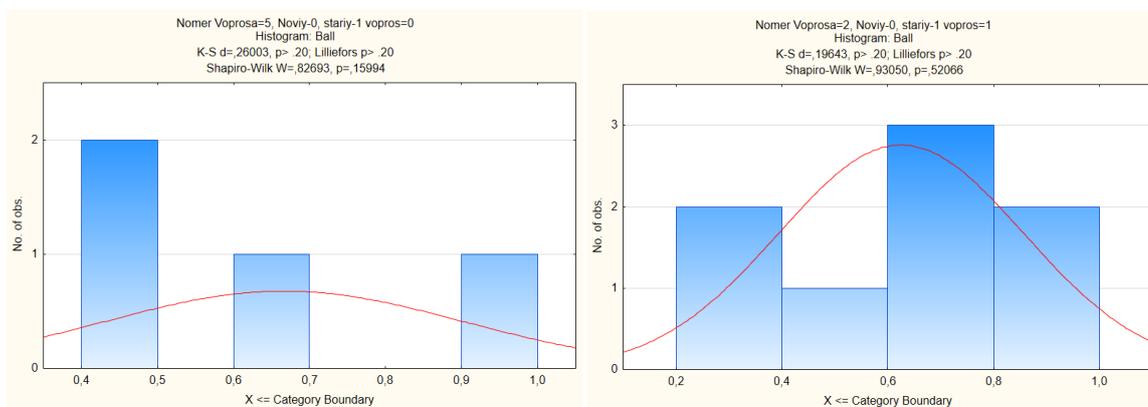


Рисунок 13 – Гистограмма частот для 5 вопроса за 2016 год и 2 вопроса за 2015 год

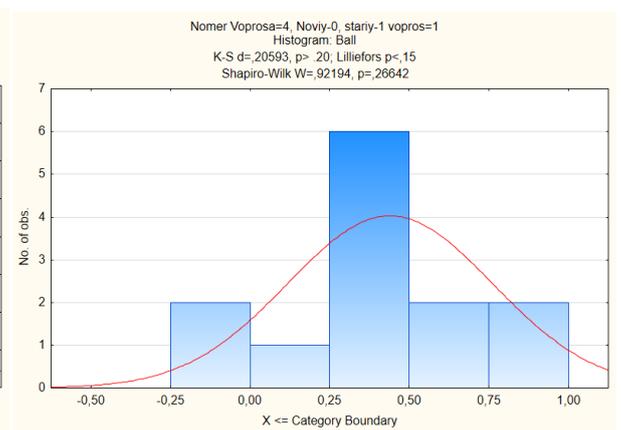
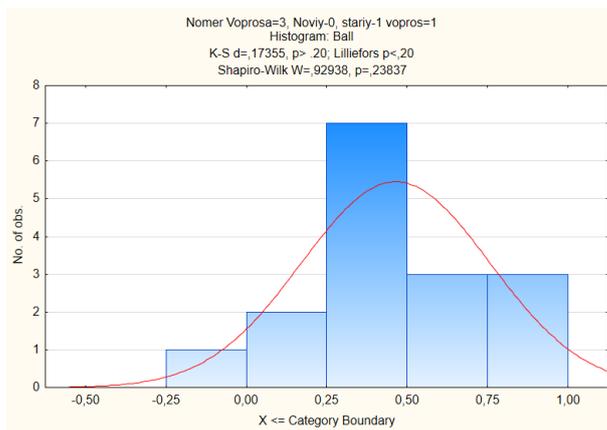


Рисунок 14 – Гистограмма частот для 3 и 4 вопросов за 2015 год

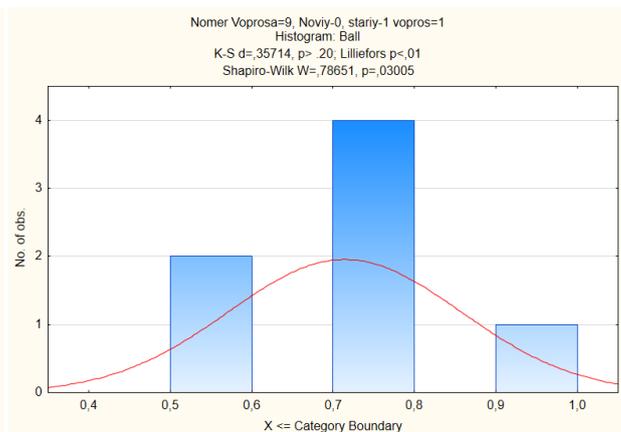
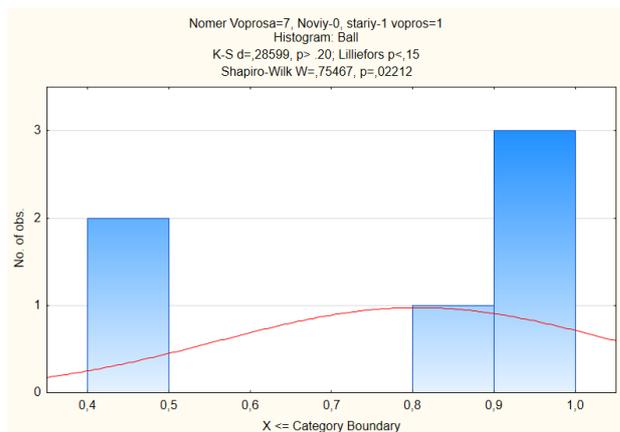


Рисунок 15 – Гистограмма частот для 7 и 9 вопросов за 2015 год

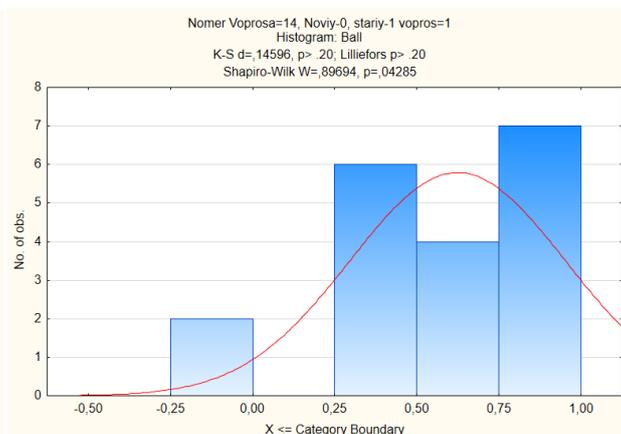
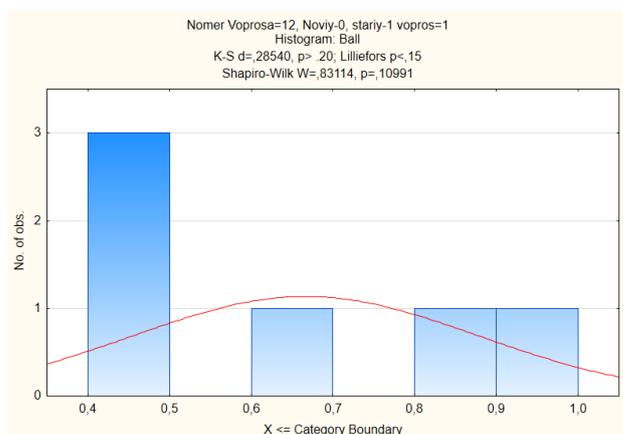


Рисунок 16 – Гистограмма частот для 12 и 14 вопросов за 2015 год

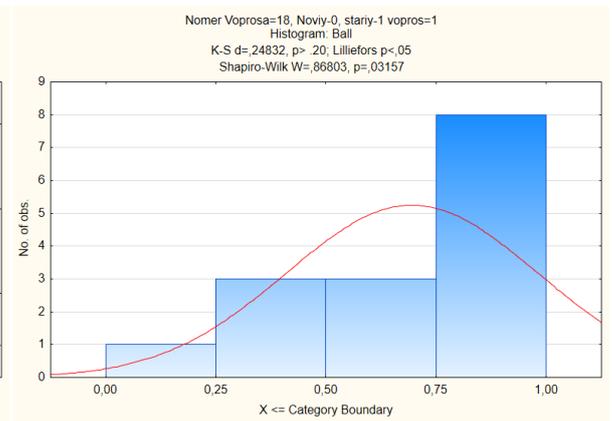
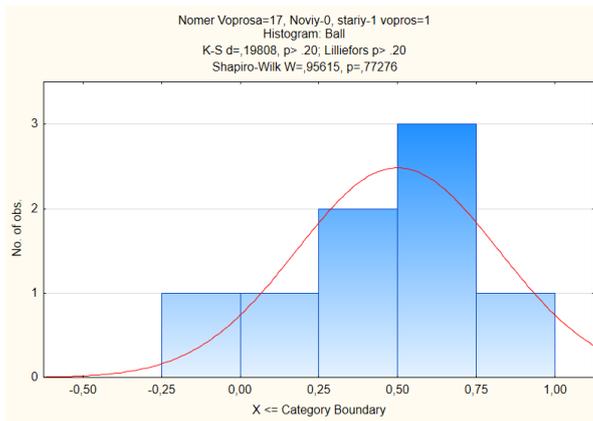


Рисунок 17 – Гистограмма частот для 17 и 18 вопросы за 2015 год

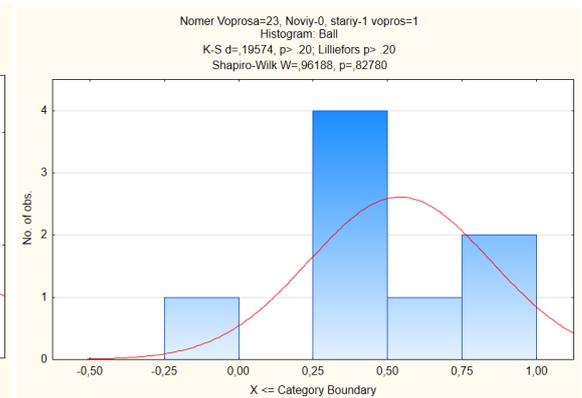
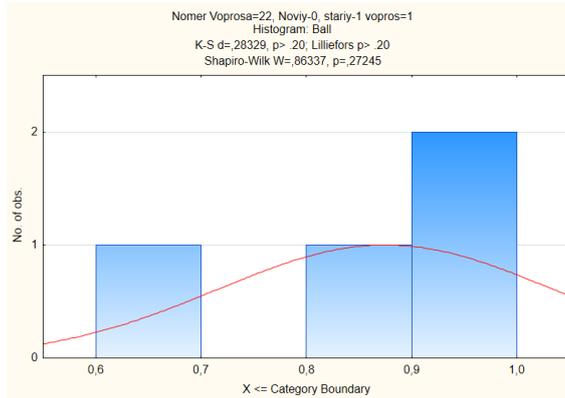


Рисунок 18 – Гистограмма частот для 22 и 23 вопросов за 2015 год

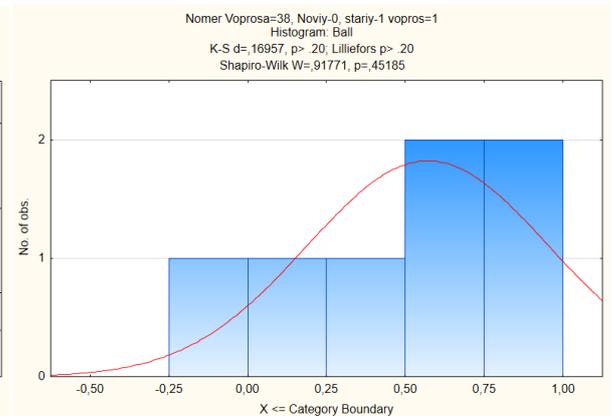
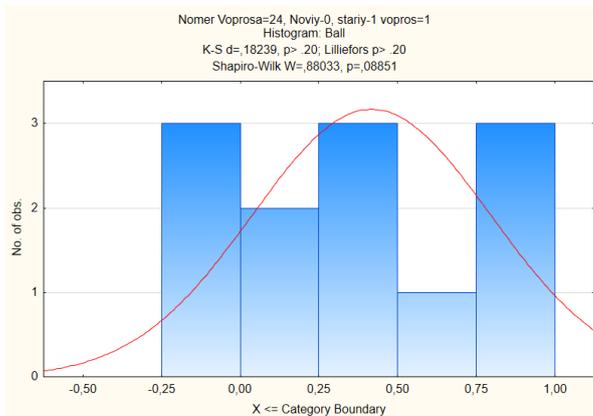


Рисунок 19 – Гистограмма частот для 24 и 38 вопросов за 2015 год

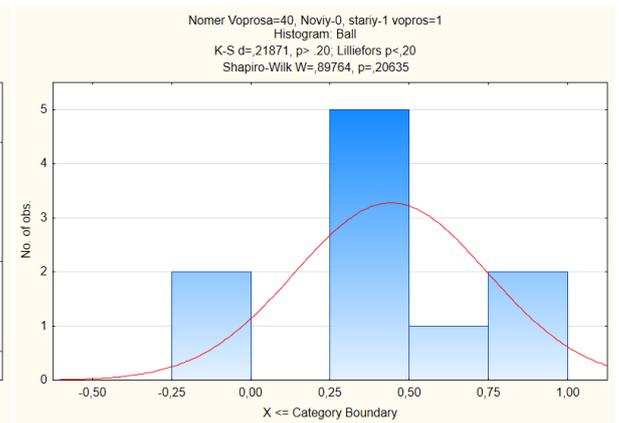
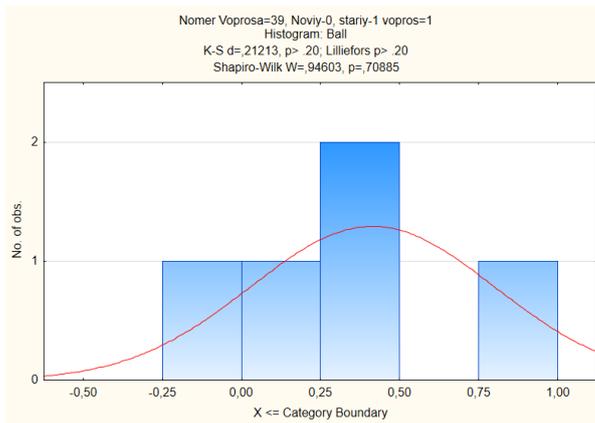


Рисунок 20 – Гистограмма частот для 39 и 40 вопросов за 2015 год

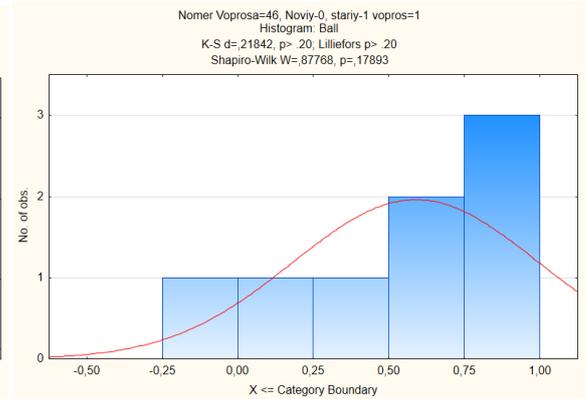
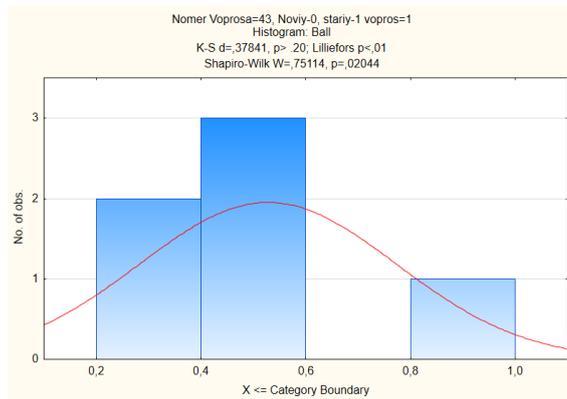


Рисунок 21 – Гистограмма частот для 43 и 46 вопросов за 2015 год

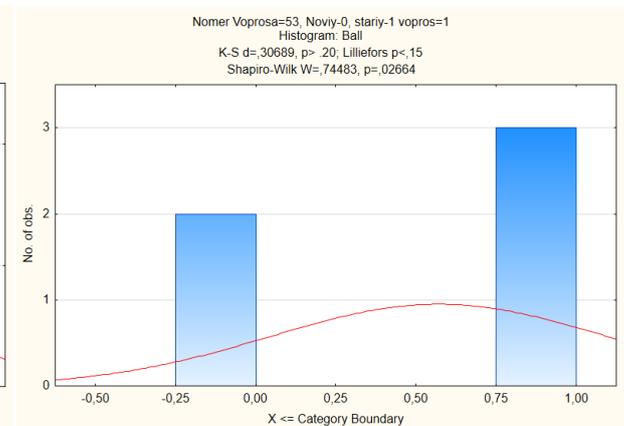
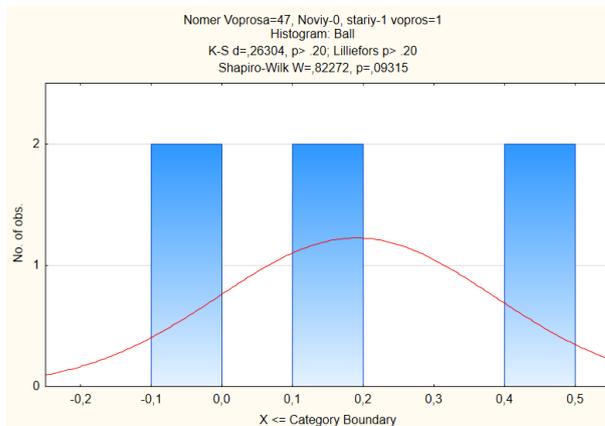


Рисунок 22 – Гистограмма частот для 47 и 53 вопросов за 2015 год

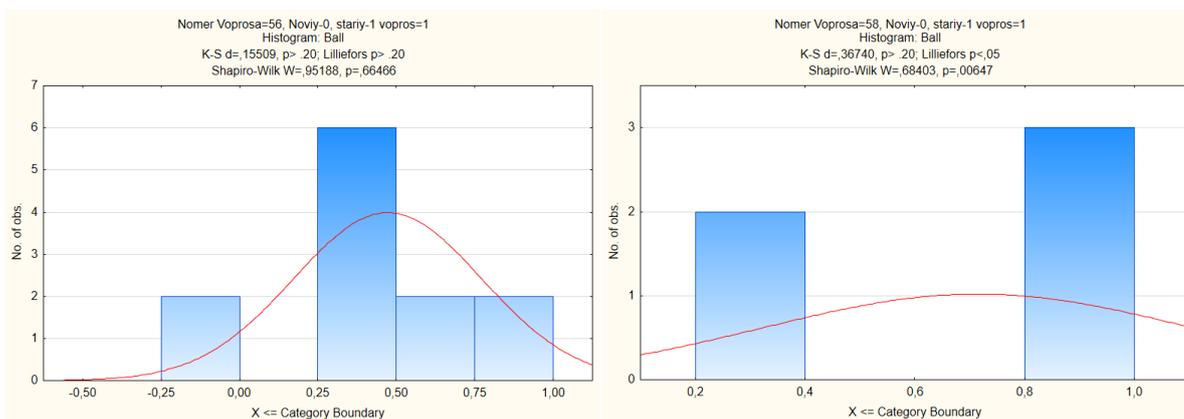


Рисунок 23 – Гистограмма частот для 56 и 58 вопросов за 2015 год

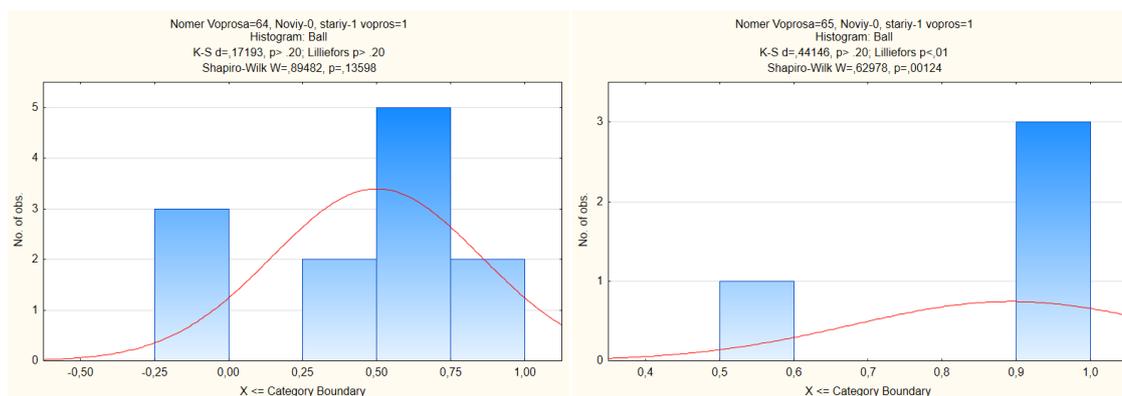


Рисунок 24 – Гистограмма частот для 64 и 65 вопросов за 2015 год

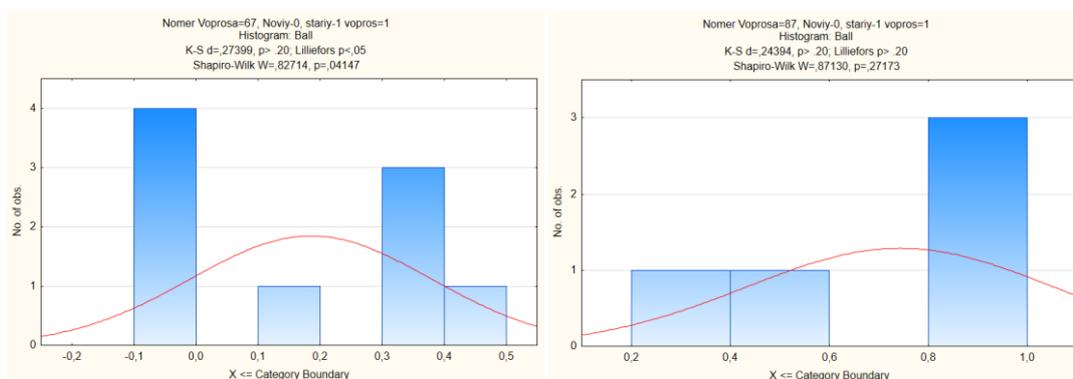


Рисунок 25 – Гистограмма частот для 67 и 87 вопросов за 2015 год

Проанализируем уровни значимости критериев Колмогорова-Смирнова, Шапиро-Уилка и Лиллиефорса. Будем искать ненормальные распределения, для этого критерии Колмогорова-Смирнова и Шапиро-Уилка должны показывать вероятность менее 5%. Среди вопросов за 2015 и 2016 год по критерию Колмогорова-Смирнова все распределения нормальные. Среди вопросов за 2016 год задания 24, 4, 46, 72, 37, 53, и за 2015 год задания 7, 9, 14, 18, 43, 53, 58, 65, 67 по критерию Шапиро-Уилка имеют ненормальные

распределения. Среди вопросов за 2016 год задания 24, 4, 46, 3, 72, 37, и за 2015 год задания 9, 18, 43, 58, 65, 67 по критерию Лиллиефорса имеют ненормальные распределения.

Далее проверим распределения на нормальность сразу за 2 года вместе (рисунок 26 – 33).

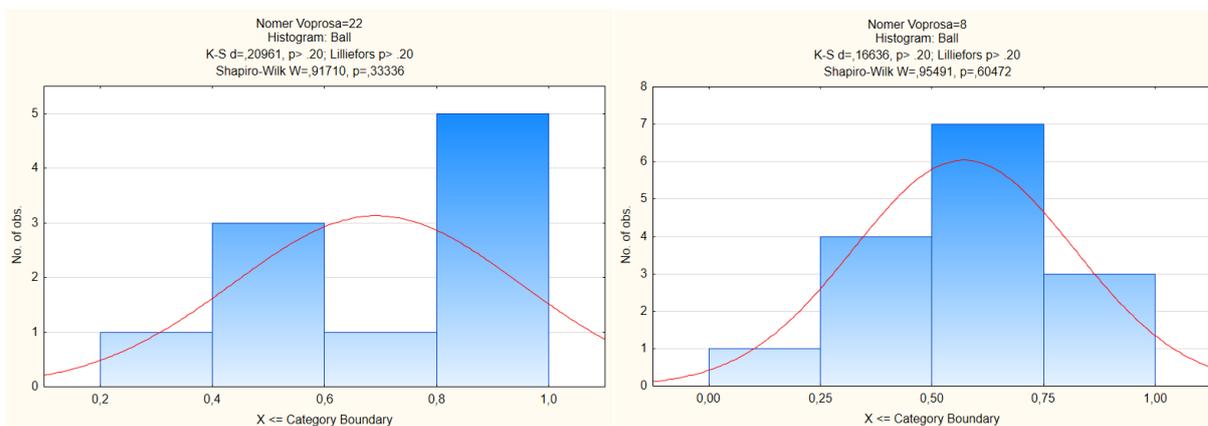


Рисунок 26 – Гистограмма частот для 22 и 8 вопросов

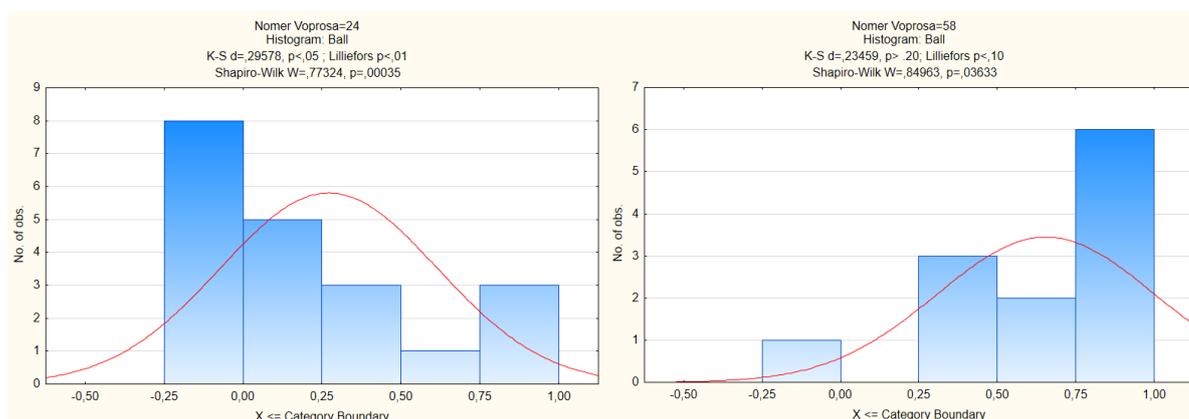


Рисунок 27 – Гистограмма частот для 24 и 58 вопросов

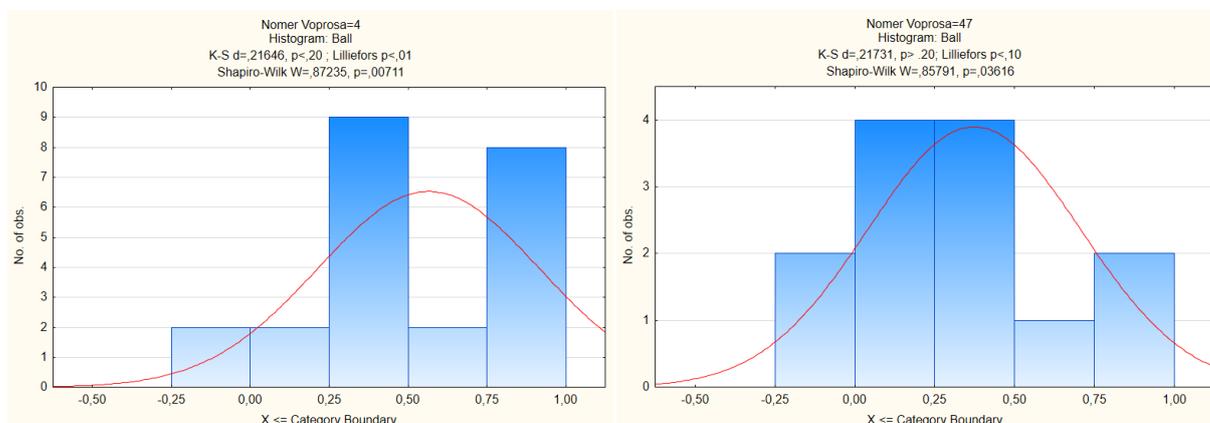


Рисунок 28 – Гистограмма частот для 4 и 47 вопросов

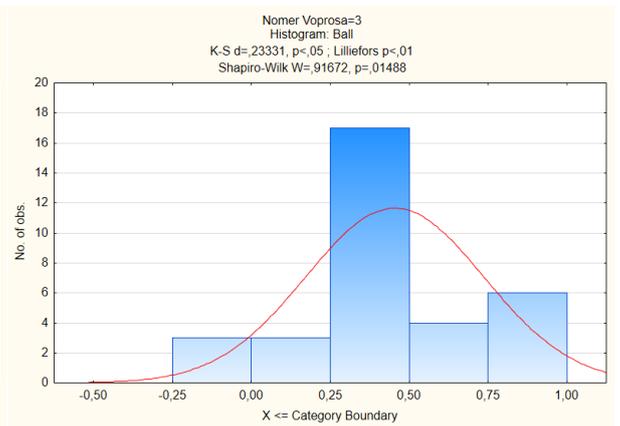
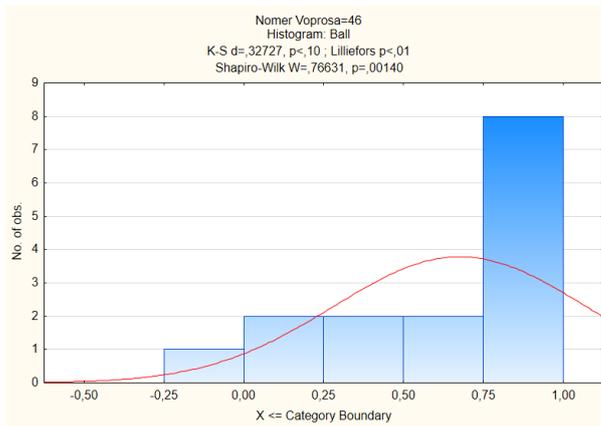


Рисунок 29 – Гистограмма частот для 46 и 3 вопросов

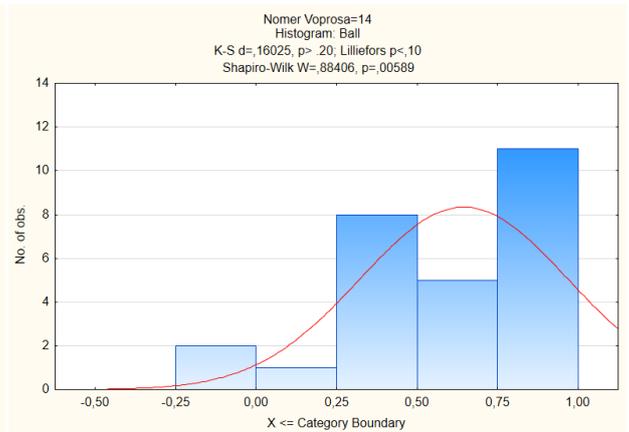
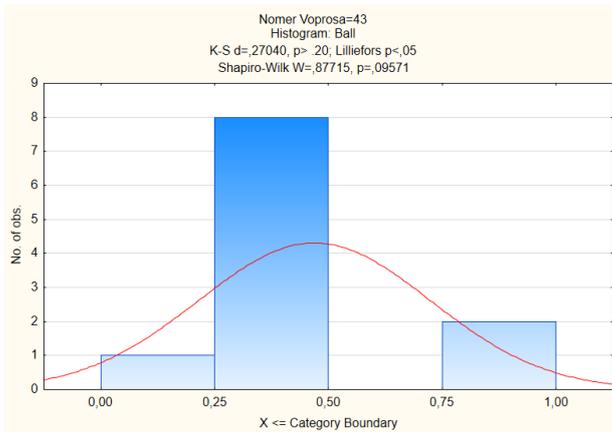


Рисунок 30 – Гистограмма частот для 43 и 14 вопросов

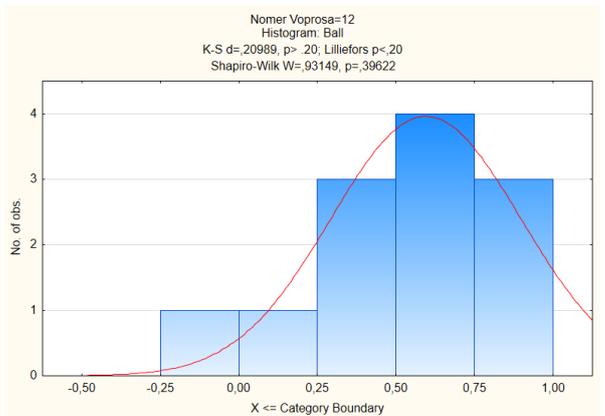
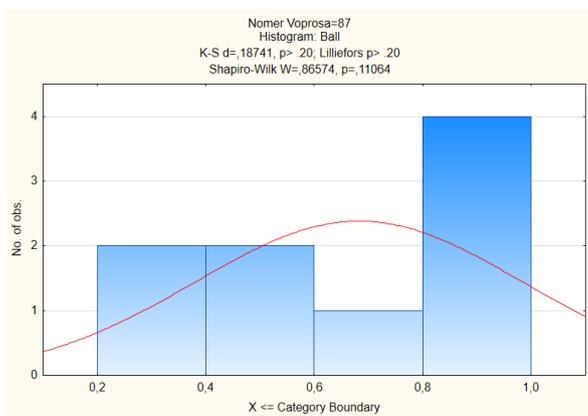


Рисунок 31 – Гистограмма частот для 87 и 12 вопросов

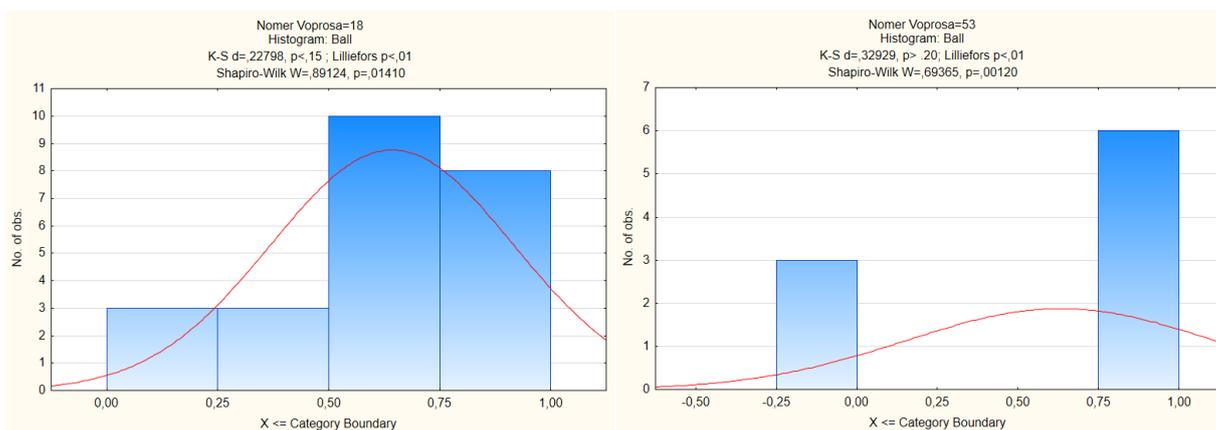


Рисунок 32 – Гистограмма частот для 18 и 53 вопросов

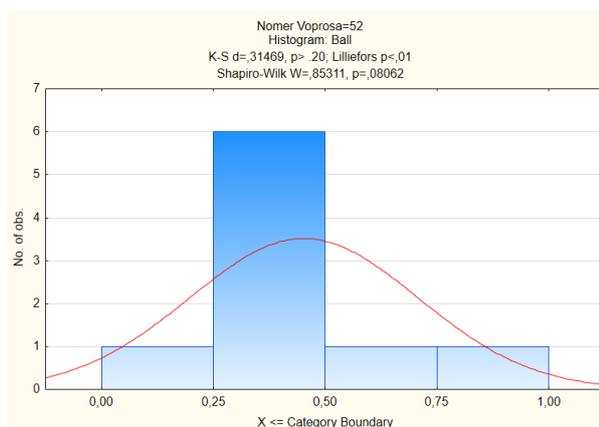


Рисунок 33 – Гистограмма частот для 52 вопроса

Проанализируем уровни значимости критериев Колмогорова-Смирнова и Шапиро-Уилка. Будем искать ненормальные распределения, для этого критерии Колмогорова-Смирнова и Шапиро-Уилка должны показывать вероятность менее 5%. По критерию Колмогорова-Смирнова задания 24, 3 имеют ненормальные распределения. Задания 24, 58, 4, 47, 46, 3, 14, 18, 53 по критерию Шапиро-Уилка имеют ненормальные распределения. По критерию Лиллиефорса задания 24, 4, 46, 3, 43, 18, 53, 52 имеют ненормальные распределения.

Как можно заметить, значительное количество распределений имеют ненормальные распределения, следовательно, можно использовать только те методы, для которых тип распределения не имеет значения.

5 КЛАСТЕРНЫЙ АНАЛИЗ

5.1 Кластерный анализ

Далее было решено разделить вопросы на 3 уровня сложности: лёгкие, средние и тяжёлые. Для этого было решено использовать кластерный анализ, а именно метод к-средних, для его использования было найдено среднее арифметическое каждого вопроса (таблица 3).

Таблица 3 – Таблица среднеарифметического баллов за каждый вопрос, с количеством ответов не менее 4

№ вопроса	Баллы за вопрос
2	0,625
3	0,464285625
4	0,439560308
7	0,8055555
8	0,591836857
9	0,714286
12	0,666666667
13	0,416666625
14	0,624060105
17	0,500000125
18	0,695238133
22	0,875
23	0,541666625
24	0,416666583
38	0,571428714
39	0,4142856
40	0,442857
43	0,527777667
44	0,62499975
45	1
46	0,58928575
47	0,190476
50	0,5833335
52	0,3714282
53	0,5714286
55	0,523809571
56	0,472222167
58	0,7142856
59	0,388889
60	0,5
64	0,500000083
65	0,89285725
66	0,306122286
67	0,185185111
87	0,7428572

Так как данные из таблицы 6 имеют один и тот же тип шкалы, то их не нужно стандартизировать. Далее к данным из таблицы 3 был применён метод к-

средних с 3 кластерами. График значений трёх кластеров показан на рисунке 34.

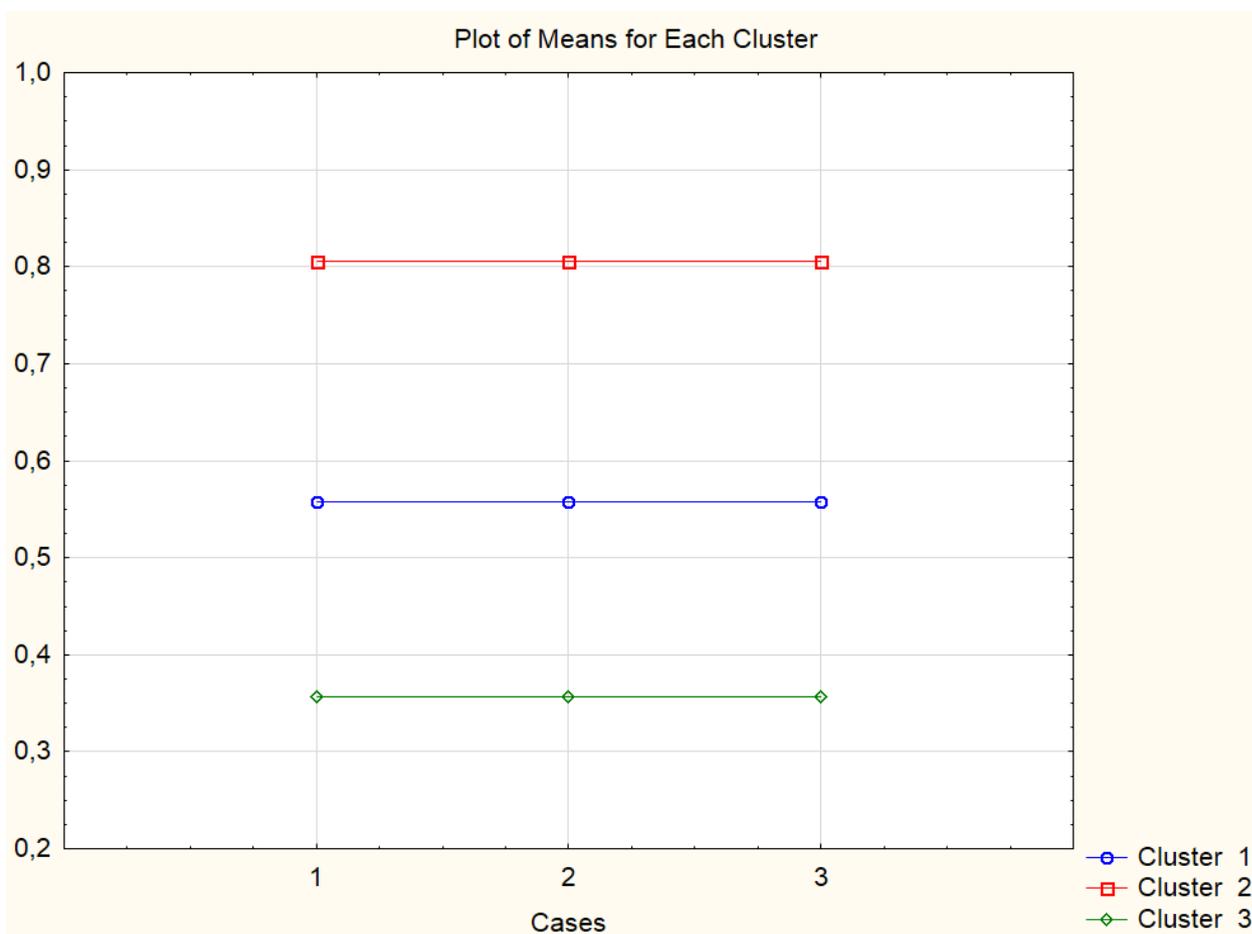


Рисунок 34 – График средних значений каждого кластера

Как видно из графика, кластер 2 состоит из лёгких вопросов, так как в него попали вопросы с высокими баллами, кластер 1 состоит из вопросов средней сложности и кластер 3 состоит из трудных вопросов.

Далее было выяснено, какие вопросы попали в какой кластер.

Кластер 1: 2, 3, 8, 12, 14, 17, 23, 38, 43, 44, 46, 50, 53, 55, 56, 60, 64 вопросы.

Кластер 2: 7, 9, 18, 22, 45, 58, 65, 87 вопросы.

Кластер 3: 4, 13, 24, 39, 40, 47, 52, 59, 66, 67 вопросы.

Для наглядности из трёх кластеров был построен график, содержащий номера вопросов и средний балл за каждый из них (рисунок 35).

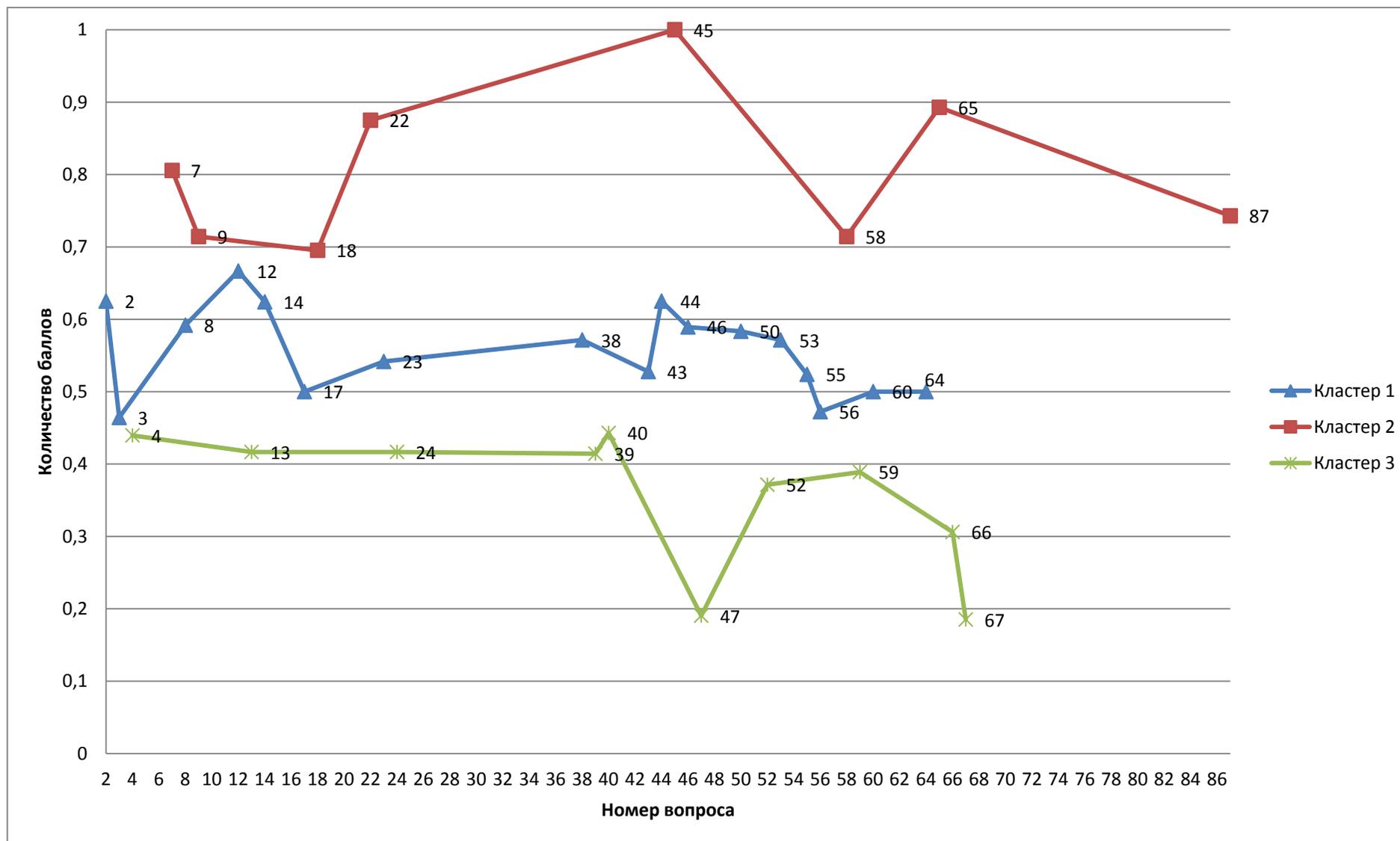


Рисунок 35 – График баллов трёх кластеров

5.2 Ранговый критерий Краскела-Уоллиса

Для того чтобы проверить, что решение разделить вопросы именно на 3 кластера было верным, нужно проверить, принадлежат ли баллы из шести тем одному и тому же распределению. Если не принадлежат одному и тому же распределению, то решение было верным, если принадлежат, то нужно разделить вопросы на другое количество кластеров. Для этого были использованы тесты Краскела-Уоллиса, Манна – Уитни и однофакторный дисперсионный анализ. Сначала был использован критерий Краскела-Уоллиса (рисунок 36).

Kruskal-Wallis ANOVA by Ranks; Ball (Voprosi bolshe 4 NEW)					
Independent (grouping) variable: Cluster					
Kruskal-Wallis test: H (2, N= 278) =52,37230 p =,0000					
Depend.: Ball	Code	Valid N	Sum of Ranks	Mean Rank	
1	1	147	21091,00	143,4762	
2	2	50	9945,00	198,9000	
3	3	81	7745,00	95,6173	

Рисунок 36 – Результаты критерия Краскела-Уоллиса

Как можно заметить, нулевую гипотезу можно принять с вероятностью $p = 0.0000$. Поскольку заданный уровень значимости много больше $\alpha = 0.05$, то нулевую гипотезу следует отвергнуть в пользу альтернативной гипотезы H_1 – влияние фактора существенное, следовательно, баллы из трёх кластеров не принадлежат одному и тому же распределению.

5.3 Критерий Манна – Уитни

Далее будем использовать критерий Манна – Уитни и сформулируем нулевую гипотезу H_0 - исходные две выборки – однородны, соответственно гипотеза H_1 утверждает, что выборки не однородны, т. е. влияние фактора значимо (рисунок 37– 39).

Mann-Whitney U Test (Voprosi bolshe 4 NEW)									
By variable Cluster									
Marked tests are significant at p <,05000									
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2
Ball	12962,00	6541,000	2084,000	-4,56718	0,000005	-4,60304	0,000004	147	50

Рисунок 37 – Результаты критерия Манна – Уитни для 1 и 2 кластеров

Mann-Whitney U Test (Voprosi bolshe 4 NEW)									
By variable Cluster									
Marked tests are significant at p <,05000									
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2
Ball	19007,00	7099,000	3778,000	4,562795	0,000005	4,587266	0,000004	147	81

Рисунок 38 – Результаты критерия Манна – Уитни для 1 и 3 кластеров

Mann-Whitney U Test (Voprosi bolshe 4 NEW)										
By variable Cluster										
Marked tests are significant at p <,05000										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	4679,000	3967,000	646,0000	6,531049	0,000000	6,581696	0,000000	50	81	0,000000

Рисунок 39 – Результаты критерия Манна – Уитни для 2 и 3 кластеров

Во всех трёх кластерах значения p-value < 0.05, и это означает, что баллы из трёх кластеров не принадлежат одному и тому же распределению.

Так как предварительный ранговый однофакторный анализ подтвердил гипотезу о значимом влиянии фактора, попробуем оценить это влияние количественно в рамках дисперсионного анализа.

5.4 Однофакторный дисперсионный анализ

Проверяем нулевую гипотезу – влияние фактора на распределение данных не существенно (рисунок 40).

Analysis of Variance (Voprosi bolshe 4 NEW)								
Marked effects are significant at p <,05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Ball	5,147886	2	2,573943	22,92651	275	0,083369	30,87406	0,000000

Рисунок 40 – Результаты дисперсионного анализа

Статистика Фишера F=30.87406 незначимо отличается от единицы с вероятностью p=0.000000, что значительно меньше уровня значимости.

Следовательно, нулевую гипотезу следует отвергнуть в пользу альтернативной гипотезы – влияние фактора существенно. И это значит, что решение разделить вопросы на 3 кластера было верным.

Теперь получим оценки эффектов обработки (рисунок 41 – 42).

Breakdown Table of Descriptive Statistics (Voprosi bolshe 4 NEW)								
N=278 (No missing data in dep. var. list)								
Cluster	Ball Means	Confidence -95,000%	Confidence +95,000%	Ball N	Ball %N	Ball Sum	Ball % Sum	Ball Std.Dev.
1	0,549887	0,501311	0,598462	147	53	80,8333	54,1986	0,297998
2	0,772381	0,699539	0,845223	50	18	38,6191	25,8940	0,256308
3	0,366549	0,302357	0,430741	81	29	29,6905	19,9074	0,290307
All Grps	0,536485	0,498898	0,574073	278	100	149,1429	100,0000	0,318358

Рисунок 41 – Влияние кластера на количество баллов

Ball Variance	Ball Std.Err.	Ball Minimum	Ball Maximum	Ball Q25	Ball Median	Ball Q75	Percentile 10,00000	Percentile 90,00000
0,088803	0,024578	0,000000	1,000000	0,333333	0,500000	0,714286	0,000000	1,000000
0,065694	0,036247	0,142857	1,000000	0,571429	0,857143	1,000000	0,285714	1,000000
0,084278	0,032256	0,000000	1,000000	0,142857	0,428571	0,500000	0,000000	0,857143
0,101352	0,019094	0,000000	1,000000	0,333333	0,500000	0,833333	0,000000	1,000000

Рисунок 42 – Влияние кластера на количество баллов

Полученные результаты (средние) свидетельствуют о существенном различии точечных характеристик для различных групп.

6 ПРОВЕРКА РАВНОЗНАЧНОСТИ СЛОЖНОСТИ ТЕМ

6.1 Анализ данных

Имеются результаты экзамена по дискретной математике. Было решено разделить вопросы на 6 тем для того, чтобы вычислить средний балл по каждой теме.

- В 1 тему “Основы теории графов” попали вопросы: 45, 52.
- Во 2 тему “Оптимизационные задачи теории графов” попали вопросы: 46, 53, 58, 65, 66, 87.
- В 3 тему “Основные понятия теории булевых функций” попали вопросы: 7, 12, 17, 38, 43, 50, 55, 59.
- В 4 тему “Нормальные формы булевых функций” попали вопросы: 2, 13, 22, 23, 56, 60, 67.
- В 5 тему “Минимизация булевых функций” попали вопросы: 3, 4, 8, 14, 18, 39, 40, 44, 64.
- В 6 тему “Не полностью определенные булевы функции и системы булевых функций” попали вопросы: 9, 24, 47.

Далее был вычислен средний балл за каждую тему. Получились следующие результаты:

- Средний балл за 1 тему составил 0,686
- Средний балл за 2 тему составил 0,636
- Средний балл за 3 тему составил 0,571
- Средний балл за 4 тему составил 0,517
- Средний балл за 5 тему составил 0,533
- Средний балл за 6 тему составил 0,440

Как можно заметить, самые высокие средние баллы получились у тем 1 и 2, а самый низкий у темы 6.

Далее был вычислен 95% доверительный интервал за каждую тему. Получились следующие результаты:

- Доверительный интервал за 1 тему составил $\pm 0,258$

- Доверительный интервал за 2 тему составил $\pm 0,068$
- Доверительный интервал за 3 тему составил $\pm 0,076$
- Доверительный интервал за 4 тему составил $\pm 0,078$
- Доверительный интервал за 5 тему составил $\pm 0,063$
- Доверительный интервал за 6 тему составил $\pm 0,141$

Для того чтобы показать распределение баллов по каждой теме более наглядно к каждой теме была применена диаграмма рассеяния (рисунок 43-45).

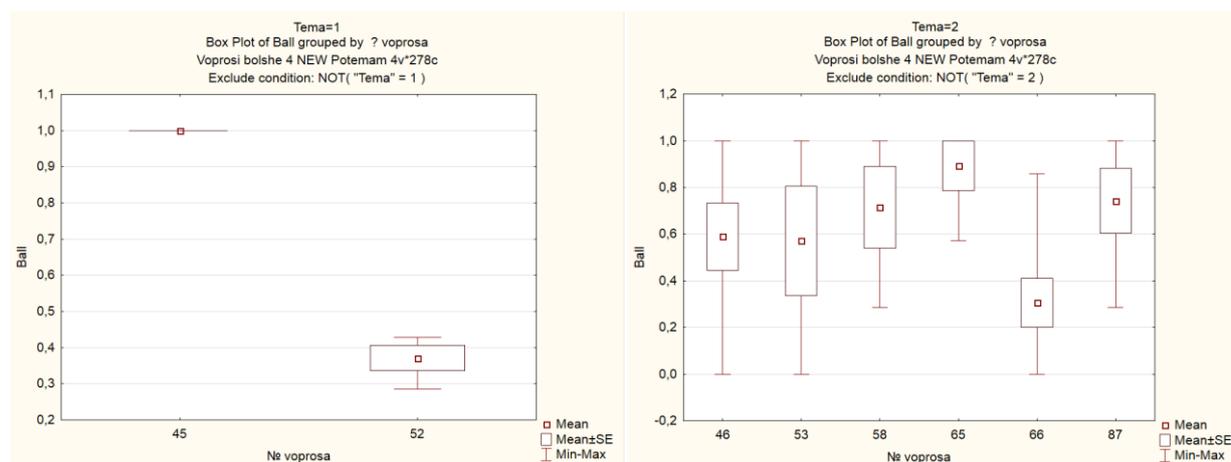


Рисунок 43 – Диаграммы рассеяния для темы 1 “Основы теории графов” и темы 2 “Оптимизационные задачи теории графов”

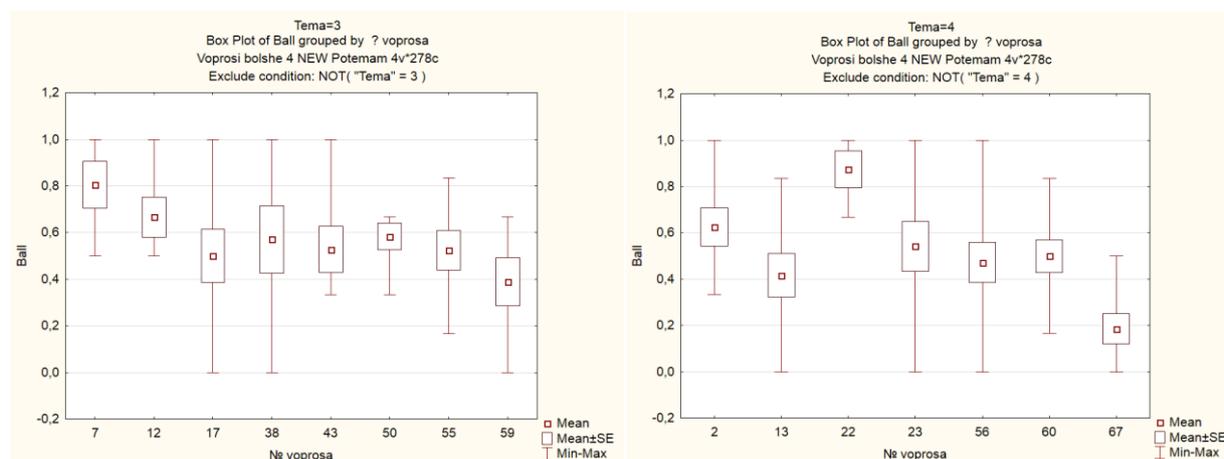


Рисунок 44 – Диаграммы рассеяния для темы 3 “Основные понятия теории булевых функций” и темы 4 “Нормальные формы булевых функций”

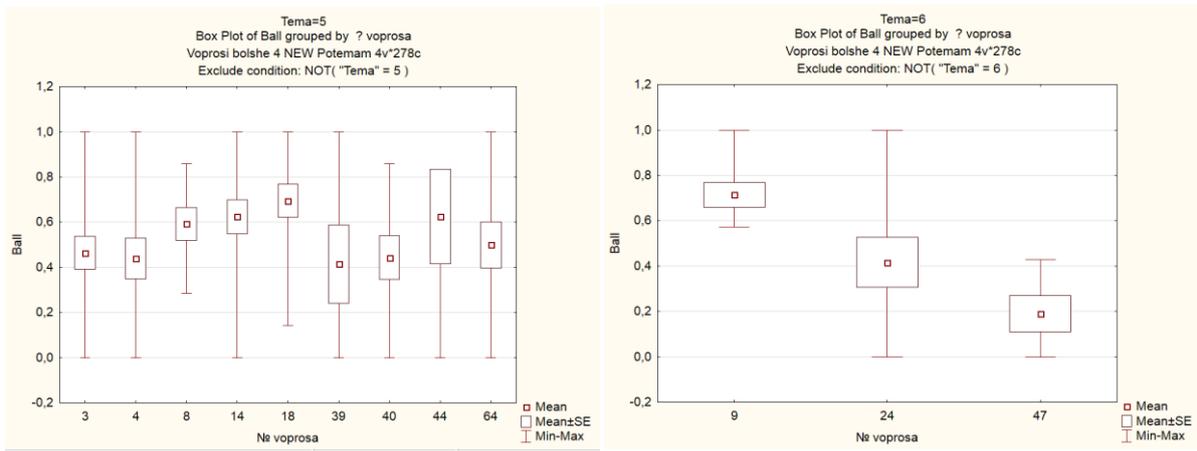


Рисунок 45 – Диаграммы рассеяния для темы 5 “Минимизация булевых функций” и темы 6 “Не полностью определенные булевы функции и системы булевых функций”

Далее было решено определить, равнозначны ли сложности тем. Для этого нужно проверить, принадлежат ли баллы из шести тем одному и тому же распределению. Если принадлежат одному и тому же распределению, то темы одинаковой сложности, если не принадлежат, то темы разной сложности.

Сформулируем нулевую гипотезу H_0 - исходные выборки однородны, т.е. принадлежат одному и тому же распределению, и соответственно гипотеза H_1 утверждает, что выборки не однородны, т.е. принадлежат разным распределениям.

6.2 Ранговый критерий Краскела-Уоллиса

Сначала для определения, принадлежат ли баллы из шести тем одному и тому же распределению, был использован ранговый критерий Краскела-Уоллиса (рисунок 46).

Kruskal-Wallis ANOVA by Ranks; Ball (Voprosi bolshe 4 NEW Potemam)					
Independent (grouping) variable: Tema					
Kruskal-Wallis test: H (5, N= 278) =5,815906 p =,3245					
Depend.:	Code	Valid N	Sum of Ranks	Mean Rank	
Ball	1	9	1445,00	160,5556	
	2	34	5290,50	155,6029	
	3	52	7718,50	148,4327	
	4	57	7192,00	126,1754	
	5	101	14149,00	140,0891	
	6	25	2986,00	119,4400	

Рисунок 46 – Ранговый критерий Краскела-Уоллиса

Нулевую гипотезу можно принять с вероятностью $p = 0.3245$. Поскольку заданный уровень значимости много больше $\alpha = 0.05$, то нулевую гипотезу следует принять – влияние фактора несущественно, следовательно, баллы принадлежат одному и тому же распределению.

6.3 Критерий Манна – Уитни

Для большей надёжности будем использовать критерий Манна – Уитни и сформулируем нулевую гипотезу H_0 - исходные две выборки – однородны, соответственно гипотеза H_1 утверждает, что выборки не однородны, т. е. влияние фактора значимо (рисунок 47-63).

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	215,5000	730,5000	135,5000	0,507519	0,611791	0,526378	0,598626	9	34	0,607287

Рисунок 47 – Сравнение 1 и 2 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	287,0000	1604,000	226,0000	0,152522	0,878775	0,155369	0,876531	9	52	0,880836

Рисунок 48 – Сравнение 1 и 3 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	354,0000	1857,000	204,0000	0,971623	0,331239	0,983069	0,325574	9	57	0,337089

Рисунок 49 – Сравнение 1 и 4 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	577,0000	5528,000	377,0000	0,839727	0,401062	0,849346	0,395689	9	101	0,407286

Рисунок 50 – Сравнение 1 и 5 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	191,5000	403,5000	78,50000	1,307706	0,190974	1,326589	0,184646	9	25	0,187952

Рисунок 51 – Сравнение 1 и 6 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	1545,000	2196,000	818,0000	0,578537	0,562902	0,586109	0,557803	34	52	0,564845

Рисунок 52 – Сравнение 2 и 3 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	1745,500	2440,500	787,5000	1,484904	0,137570	1,498096	0,134109	34	57	0,137051

Рисунок 53 – Сравнение 2 и 4 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	2522,500	6657,500	1506,500	1,064487	0,287109	1,075456	0,282171	34	101	0,287411

Рисунок 54 – Сравнение 2 и 5 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	1127,000	643,0000	318,0000	1,633635	0,102336	1,657817	0,097355	34	25	0,102645

Рисунок 55 – Сравнение 2 и 6 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < .05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	3115,500	2879,500	1226,500	1,547018	0,121860	1,575914	0,115047	52	57	0,121442

Рисунок 56 – Сравнение 3 и 4 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)									
By variable Tema									
Marked tests are significant at $p < ,05000$									
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2
Ball	4152,500	7628,500	2477,500	0,570072	0,568629	0,573175	0,566527	52	101

Рисунок 57 – Сравнение 3 и 5 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < ,05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	2162,500	840,5000	515,5000	1,457728	0,144917	1,471585	0,141134	52	25	0,144240

Рисунок 58 – Сравнение 3 и 6 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)									
By variable Tema									
Marked tests are significant at $p < ,05000$									
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2
Ball	4226,500	8334,500	2573,500	-1,10251	0,270242	-1,10788	0,267913	57	101

Рисунок 59 – Сравнение 4 и 5 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < ,05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	2400,500	1002,500	677,5000	0,347507	0,728211	0,350169	0,726212	57	25	0,726168

Рисунок 60 – Сравнение 4 и 6 тем

Mann-Whitney U Test (Voprosi bolshe 4 NEW Potemam)										
By variable Tema										
Marked tests are significant at $p < ,05000$										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value	Z adjusted	p-value	Valid N Group 1	Valid N Group 2	2*1sided exact p
Ball	6604,500	1396,500	1071,500	1,165341	0,243882	1,177143	0,239140	101	25	0,244291

Рисунок 61 – Сравнение 5 и 6 тем

Multiple Comparisons z' values; Ball (Voprosi bolshe 4 NEW Po						
Independent (grouping) variable: Tema						
Kruskal-Wallis test: $H(5, N=278) = 5,815906$ $p = ,3245$						
Depend.:	1	2	3	4	5	6
Ball	R:160,56	R:155,60	R:148,43	R:126,18	R:140,09	R:119,44
1		0,164334	0,417666	1,192230	0,731803	1,315601
2	0,164334		0,404382	1,689179	0,973237	1,707314
3	0,417666	0,404382		1,443655	0,608045	1,481771
4	1,192230	1,689179	1,443655		1,044666	0,349247
5	0,731803	0,973237	0,608045	1,044666		1,149775
6	1,315601	1,707314	1,481771	0,349247	1,149775	

Рисунок 62 – Множественное сравнение Z' значений тем

Multiple Comparisons p values (2-tailed); Ball (Voprosi bolshe 4 Independent (grouping) variable: Tema Kruskal-Wallis test: H (5, N= 278) =5,815906 p =,3245						
Depend.: Ball	1	2	3	4	5	6
	R:160,56	R:155,60	R:148,43	R:126,18	R:140,09	R:119,44
1		1,000000	1,000000	1,000000	1,000000	1,000000
2	1,000000		1,000000	1,000000	1,000000	1,000000
3	1,000000	1,000000		1,000000	1,000000	1,000000
4	1,000000	1,000000	1,000000		1,000000	1,000000
5	1,000000	1,000000	1,000000	1,000000		1,000000
6	1,000000	1,000000	1,000000	1,000000	1,000000	

Рисунок 63 – Множественное сравнение вероятностей тем

Результаты сравнения всех шести тем показали значения p-value > 0.05, и это означает, что баллы из шести тем принадлежат одному и тому же распределению.

6.4 Группировка тем

Так как было выяснено, что темы принадлежат одному распределению и, следовательно, их сложности равнозначны, то теперь будем группировать темы для того, чтобы найти такое сочетание тем, при котором они будут принадлежать разным распределениям.

Объединим темы 1 и 2, и 3 и 5. Объединённые темы 1 и 2 будут называться 1, а 3 и 5 будут называться 3. (рисунок 64-66).

Kruskal-Wallis ANOVA by Ranks; Ball (Voprosi bolshe 4 NEW Potemam New) Independent (grouping) variable: TemaNew Kruskal-Wallis test: H (3, N= 278) =5,414696 p =,1438				
Depend.: Ball	Code	Valid N	Sum of Ranks	Mean Rank
1	1	43	6735,50	156,6395
3	3	153	21867,50	142,9248
4	4	57	7192,00	126,1754
6	6	25	2986,00	119,4400

Рисунок 64 – Ранговый критерий Краскела-Уоллиса

Multiple Comparisons z' values; Ball (Voprosi bolshe 4 NEW Potemam New) Independent (grouping) variable: TemaNew Kruskal-Wallis test: H (3, N= 278) =5,414696 p =,1438				
Depend.: Ball	1	3	4	6
	R:156,64	R:142,92	R:126,18	R:119,44
1		0,988335	1,875971	1,839727
3	0,988335		1,342576	1,354127
4	1,875971	1,342576		0,349247
6	1,839727	1,354127	0,349247	

Рисунок 65 – Множественное сравнение Z' значений тем

Multiple Comparisons p values (2-tailed); Ball (Voprosi bolshe 4 NEW Potemam New) Independent (grouping) variable: TemaNew Kruskal-Wallis test: H (3, N= 278) =5,414696 p =,1438				
Depend.: Ball	1	3	4	6
	R:156,64	R:142,92	R:126,18	R:119,44
1		1,000000	0,363956	0,394850
3	1,000000		1,000000	1,000000
4	0,363956	1,000000		1,000000
6	0,394850	1,000000	1,000000	

Рисунок 66 – Множественное сравнение вероятностей тем

Как можно заметить, темы всё ещё принадлежат одному распределению.

Попробуем теперь объединить ещё и темы 4 и 6. Объединённые темы 4 и 6 будут называться 4 (рисунок 67-69).

Kruskal-Wallis ANOVA by Ranks; Ball (Voprosi bolshe 4 NEW Potemam New) Independent (grouping) variable: TemaNewNew Kruskal-Wallis test: H (2, N= 278) =5,291344 p =,0710				
Depend.: Ball	Code	Valid N	Sum of Ranks	Mean Rank
1	1	43	6735,50	156,6395
3	3	153	21867,50	142,9248
4	4	82	10178,00	124,1220

Рисунок 67 – Ранговый критерий Краскела-Уоллиса

Multiple Comparisons z' values; Ball (Voprosi bolshe 4 NEW Potemam New) Independent (grouping) variable: TemaNewNew Kruskal-Wallis test: H (2, N= 278) =5,291344 p =,0710			
Depend.: Ball	1	3	4
	R:156,64	R:142,92	R:124,12
1		0,988335	2,148179
3	0,988335		1,708871
4	2,148179	1,708871	

Рисунок 68 – Множественное сравнение Z' значений тем

Multiple Comparisons p values (2-tailed); Ball (Voprosi bolshe 4 NEW Potemam New) Independent (grouping) variable: TemaNewNew Kruskal-Wallis test: H (2, N= 278) =5,291344 p =,0710			
Depend.: Ball	1	3	4
	R:156,64	R:142,92	R:124,12
1		0,968965	0,095099
3	0,968965		0,262424
4	0,095099	0,262424	

Рисунок 69 – Множественное сравнение вероятностей тем

Как можно заметить, темы всё ещё принадлежат одному распределению.

Попробуем теперь объединить ещё и ранее уже объединённые темы 1 и 3. Объединённые темы 1 и 3 будут называться 3(рисунок 70-72).

Kruskal-Wallis ANOVA by Ranks; Ball (Voprosi bolshe 4 NEW Potemam New)				
Independent (grouping) variable: TemaNewNewNew				
Kruskal-Wallis test: H (1, N= 278) =4,303491 p =,0380				
Depend.: Ball	Code	Valid N	Sum of Ranks	Mean Rank
3	3	196	28603,00	145,9337
4	4	82	10178,00	124,1220

Рисунок 70 – Ранговый критерий Краскела-Уоллиса

Multiple Comparisons z' values; Ball (Voprosi bolshe 4 NEW Potemam New)				
Independent (grouping) variable: TemaNewNewNew				
Kruskal-Wallis test: H (1, N= 278) =4,303491 p =,0380				
Depend.: Ball	3	4		
	R:145,93	R:124,12		
3		2,062855		
4	2,062855			

Рисунок 71 – Множественное сравнение Z' значений тем

Multiple Comparisons p values (2-tailed); Ball (Voprosi bolshe 4 NEW Potemam New)				
Independent (grouping) variable: TemaNewNewNew				
Kruskal-Wallis test: H (1, N= 278) =4,303491 p =,0380				
Depend.: Ball	3	4		
	R:145,93	R:124,12		
3		0,039126		
4	0,039126			

Рисунок 72 – Множественное сравнение вероятностей тем

Теперь удалось получить выборки, которые относятся к разным распределениям. Следовательно, темы 1, 2, 3, 5 относятся к одному распределению, а темы 4 и 6 относятся к другому распределению.

6.5 Однофакторный дисперсионный анализ

Далее для большей надежности принадлежат ли темы 1, 2, 3, 5 и 4, 6 разным распределениям был применён однофакторный дисперсионный анализ (рисунок 73-74).

TemaNewNewNew=3								
Analysis of Variance (Voprosi bolshe 4 NEW Potemam New)								
Marked effects are significant at p < ,05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Ball	3,789626	24	0,157901	16,16205	171	0,094515	1,670647	0,032531

Рисунок 73 – Результаты дисперсионного анализа для тем 1, 2, 3, 5

TemaNewNewNew=4								
Analysis of Variance (Voprosi bolshe 4 NEW Potemam New)								
Marked effects are significant at p < ,05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Ball	2,570015	9	0,285557	5,057807	72	0,070247	4,065027	0,000301

Рисунок 74 – Результаты дисперсионного анализа для тем 4, 6

Статистика Фишера $F=1.67$ и $F=4.06$ незначимо отличается от единицы с вероятностью $p=0.03$ и $p=0.00$, что значительно меньше уровня значимости. Следовательно, нулевую гипотезу следует отвергнуть в пользу альтернативной гипотезы – влияние фактора существенно. Подводя итоги можно сказать, что все 6 тем равнозначны по сложности, но всё же темы 4 и 6 для студентов являются чуть более сложными, чем темы 1, 2, 3, 5. Так же было сопоставлено, к какой теме принадлежат вопросы, из каких кластеров.

- Тема 1 “Основы теории графов” содержит в себе вопрос из кластера с лёгкими вопросами – вопрос 45, не содержит вопросы из кластера с вопросами средней сложности, и содержит вопрос из кластера с трудными вопросами – вопрос 52.
- Тема 2 “Оптимизационные задачи теории графов” содержит в себе вопросы из кластера с лёгкими вопросами – вопросы 58, 65, 87, содержит вопросы из кластера с вопросами средней сложности – 46, 53, содержит вопрос из кластера с трудными вопросами – 66.
- Тема 3 “Основные понятия теории булевых функций” содержит в себе вопрос из кластера с лёгкими вопросами – вопрос 7, содержит вопросы из кластера с вопросами средней сложности – 12, 17, 38, 43, 50, 55, содержит вопрос из кластера с трудными вопросами – 59.
- Тема 4 “Нормальные формы булевых функций” содержит в себе вопрос из кластера с лёгкими вопросами – вопрос 22, содержит вопросы из кластера с вопросами средней сложности – 2, 23, 56, 60, содержит вопросы из кластера с трудными вопросами – 13, 67.
- Тема 5 “Минимизация булевых функций” содержит в себе вопросы из кластера с лёгкими вопросами – вопрос 45, 18, содержит вопросы из

кластера с вопросами средней сложности – 3, 8, 14, 44, 64, содержит вопросы из кластера с трудными вопросами – 4, 39, 40.

- Тема 6 “Не полностью определенные булевы функции и системы булевых функций” не содержит в себе вопросы из кластера с лёгкими вопросами, содержит вопрос из кластера с вопросами средней сложности – 9, содержит вопросы из кластера с трудными вопросами – 24, 47.

Подводя итоги распределения сложности вопросов по темам можно сделать следующие выводы:

Тема 1 содержит 1 легкий, 0 средних, 1 сложный вопрос.

Тема 2 содержит 3 легких, 2 средних, 1 сложный вопрос.

Тема 3 содержит 1 легкий, 6 средних, 1 сложный вопрос.

Тема 4 содержит 1 легкий, 4 средних, 2 сложных вопроса.

Тема 5 содержит 2 легких, 5 средних, 3 сложных вопроса.

Тема 6 содержит 0 легких, 1 средний, 2 сложных вопроса.

РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

Целью данного раздела является комплексный анализ и описание финансово-экономических аспектов выполненной работы. Необходимо оценить полные денежные затраты на проект, а также дать приближенную экономическую оценку результатов ее внедрения. Это, в свою очередь, позволит оценить экономическую целесообразность осуществления работы с помощью традиционных показателей инвестиций.

9.1 Организация и планирование работ

В данном пункте определяется перечень проводимых работ, их исполнители и оптимальная продолжительность выполнения. Для дальнейшего определения продолжительности этапов работ и их трудоемкости по каждому исполнителю, а также построения линейного графика реализации проекта этапы работы хронологически упорядочены и сведены в таблицу 9, где НР – научный руководитель, И – инженер (исполнитель проекта).

Таблица 9 – Перечень работ и продолжительность их выполнения

Этапы работы	Исполнители	Загрузка исполнителей
Постановка целей и задач, получение исходных данных	НР, И	НР – 100 % И – 10 %
Составление и утверждение технического задания	НР, И	НР – 60 % И – 40 %
Подбор и изучение материалов по тематике	НР, И	НР – 10 % И – 100 %
Разработка календарного плана	НР, И	НР – 50 % И – 50 %
Изучение математических методов и вспомогательного программного обеспечения	НР, И	НР – 5 % И – 100 %
Выполнение проверки данных на нормальность распределения	И	И – 100 %
Выполнение кластерного анализа	И	И – 100 %
Проверка равнозначности сложности тем	И	И – 100 %
Сравнение данных экзамена за 2015 и 2016 год	И	И – 100 %
Выполнение регрессионного анализа	И	И – 100 %

Анализ и обсуждение полученных результатов	НР, И	НР – 80 % И – 100 %
Оформление пояснительной записки	И	И – 100%

9.1.1 Продолжительность этапов работ

Расчет продолжительности выполнения этапов работ осуществлен опытно-статистическим методом с применением экспертного способа.

Ожидаемые значения продолжительности выполнения работ $t_{ож}$ определяются по формуле:

$$t_{ож} = \frac{3 \cdot t_{min} + 2 \cdot t_{max}}{5},$$

где t_{min} – минимальная продолжительность выполнения этапа, дн.;

t_{max} – максимальная продолжительность выполнения этапа, дн.

Расчет длительности выполнения каждого этапа работы в рабочих днях выполняется по формуле:

$$T_{РД} = \frac{t_{ож}}{K_{ВН}} \cdot K_{Д},$$

где $t_{ож}$ – ожидаемая продолжительность этапа, дн.;

$K_{ВН}$ – коэффициент выполнения этапа, учитывающий влияние внешних факторов на соблюдение предварительно определенных длительностей. $K_{ВН} = 1$;

$K_{Д}$ – коэффициент, учитывающий дополнительное время на компенсацию непредвиденных задержек и согласование работ. $K_{Д} = 1,2$.

Для перевода значений длительности этапа в рабочих днях к их аналогам в календарных днях используется формула:

$$T_{кд} = T_{РД} \cdot T_{К},$$

где $T_{РД}$ – продолжительность выполнения этапа в рабочих днях;

$T_{К}$ – коэффициент календарности.

Значение $T_{К}$ определяется следующим образом:

$$T_{К} = \frac{T_{КАЛ}}{T_{КАЛ} - T_{ВД}},$$

где $T_{КАЛ}$ – календарные дни;

$T_{ВД}$ – выходные и праздничные дни;

$$T_K = \frac{365}{365 - 66} = 1,22 .$$

Данное значение применимо для шестидневной рабочей недели. Для пятидневной рабочей недели $T_{ВД} = 118$, следовательно $T_K = 1,48$.

Результаты выполненных расчетов по определению продолжительности этапов работ и их трудоемкости по исполнителям приведены в таблице 10.

Таблица 10 – Трудозатраты на выполнение проекта

Этап	Исполнители	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.-дн.			
					$T_{РД}$		$T_{КД}$	
		t_{min}	t_{max}	$t_{ож}$	НР	И	НР	И
1	2	3	4	5	6	7	8	9
1. Постановка целей и задач, получение исходных данных	НР, И	2	4	2,8	3,36	0,34	4,97	0,41
2. Составление и утверждение технического задания	НР, И	4	6	4,8	3,46	2,3	5,11	2,81
3. Подбор и изучение материалов по тематике	НР, И	8	10	8,8	1,06	10,56	1,29	12,88
4. Разработка календарного плана	НР, И	2	4	2,8	1,68	1,68	2,49	2,05
5. Изучение математических методов и вспомогательного программного обеспечения	НР, И	7	10	8,2	0,49	9,84	0,73	12
6. Выполнение проверки данных на нормальность распределения	И	5	7	5,8	–	6,96	–	8,49
7. Выполнение кластерного анализа	И	7	9	7,8	–	9,36	–	11,42
8. Проверка равнозначности сложности тем	И	10	13	11,2	–	13,44	–	16,4
9. Сравнение данных экзамена за 2015 и 2016 год	И	5	7	5,8	–	6,96	–	10,30
10. Выполнение регрессионного анализа	И	5	6	5,4	–	6,48	–	9,59

11. Анализ и обсуждение полученных результатов	НР, И	2	5	3,2	3,07	3,84	4,54	4,68
12. Оформление пояснительной записки	И	14	20	16,4	–	19,68	–	24,01
Итого:				83	13,12	91,44	19,13	115,04

Таблица 11 – Линейный график работ

Этап	НР	И	Февраль			Март			Апрель			Май		
			10	20	30	40	50	60	70	80	90	100	110	120
1	4,97	0,41	■											
2	5,11	2,81		■										
3	1,29	12,88			■									
4	2,49	2,05					■							
5	0,73	12						■						
6	–	8,49							■					
7	–	11,42								■				
8	–	16,4									■			
9	–	10,30										■		
10	–	9,59											■	
11	4,54	4,68												■
12	–	24,01												■

НР – ■; И – ■.

9.1.2 Расчет накопления готовности проекта

В данном пункте выполняется оценка текущих результатов работы над проектом. Величина накопления готовности работы показывает на сколько процентов, по окончании текущего этапа, выполнен общий объем по проекту.

Введем обозначения:

$TP_{\text{общ}}$ – общая трудоемкость проекта;

TP_i (TP_k) – трудоемкость i -го (k -го) этапа проекта, $i = \overline{1, I}$;

TP_i^H – накопленная трудоемкость i -го этапа проекта по его завершении;

TP_{ij} (TP_{kj}) – трудоемкость работ, выполняемых j -м участником на i -м этапе, здесь $j = \overline{1, m}$ – индекс исполнителя, в нашем примере $m = 2$.

Степень готовности определяется формулой:

$$CG_i = \frac{TP_i^H}{TP_{\text{общ.}}} = \frac{\sum_{k=1}^i TP_k}{TP_{\text{общ.}}} = \frac{\sum_{k=1}^i \sum_{j=1}^m TP_{km}}{\sum_{k=1}^i \sum_{j=1}^m TP_{km}}.$$

Результаты вычислений отражены в таблице 12.

Таблица 12 – Нарастание технической готовности работы и удельный вес каждого этапа

Этапы работы	$TP_i, \%$	$CG_i, \%$
Постановка целей и задач, получение исходных данных	3,53	3,53
Составление и утверждение технического задания	5,50	9,04
Подбор и изучение материалов по тематике	11,11	20,16
Разработка календарного плана	3,21	23,37
Изучение математических методов и вспомогательного программного обеспечения	9,87	33,25
Выполнение проверки данных на нормальность распределения	6,65	39,91
Выполнение кластерного анализа	8,95	48,86
Проверка равнозначности сложности тем	12,85	61,71
Сравнение данных экзамена за 2015 и 2016 год	6,65	68,37
Выполнение регрессионного анализа	6,19	74,56
Анализ и обсуждение полученных результатов	6,60	81,17
Оформление пояснительной записки	18,82	100,00

9.2 Расчет сметы затрат на выполнение проекта

К затратам на создание проекта относится величина всех расходов, которые необходимы для реализации комплекса работ, составляющих содержание разработки.

9.2.1 Расчет затрат на материалы

К данной статье расходов относится стоимость материалов, покупных изделий и других материальных ценностей, расходуемых в процессе выполнения работ. Цена материальных ресурсов определяется по соответствующим им ценникам. Сюда же включаются расходы на совершение сделки купли-продажи.

Кроме того, данная статья включает в себя транспортно-заготовительные расходы, которые связаны с транспортировкой от поставщика к потребителю, хранением и прочими процессами, обеспечивающими движение материальных ресурсов от поставщиков к потребителю. ТРЗ составляют 5 % от отпускной цены материалов.

Таблица 13 – Расчет затрат на материалы

Наименование материалов	Цена за ед., руб.	Кол-во	Сумма, руб.
Бумага для принтера формата А4	229	1 уп.	299
Картридж для принтера	1399	1 шт.	1399
Канцелярские принадлежности	120	1 комп.	120
Итого			1818
ТЗР 5 %			90,9
Итого с ТЗР			1908,9

9.2.2 Расчет заработной платы

К данной статье расходов относится заработная плата научного руководителя и исполнителя проекта, а также премии, которые входят в фонд заработной платы. Расчет основной заработной платы выполняется на основе трудоемкости выполнения каждого этапа работы и величины месячного оклада.

Среднедневная тарифная заработная плата ($ЗП_{дн-т}$) рассчитывается по формуле:

$$ЗП_{дн-т} = MO/24,92.$$

Расчеты затрат на основную заработную плату приведены в таблице 14. При расчете учитывалось, что в году 299 рабочих дней и, следовательно, в месяце 24,92 рабочих дня. Затраты времени на выполнение работы по каждому исполнителю брались из таблицы 10. Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется следующий ряд коэффициентов: $K_{ГП} = 1,1$; $K_{доп.ЗП} = 1,188$; $K_p = 1,3$. Таким образом, для перехода от базовой суммы заработка исполнителя, к соответствующему полному заработку необходимо первую умножить на интегральный коэффициент $K_{и} = 1,699$. Вышеуказанное значение $K_{доп.ЗП}$ применяется при шестидневной рабочей неделе, при пятидневной оно равно 1,113, соответственно в этом случае $K_{и} = 1,62$.

Таблица 14 Затраты на заработную плату

Исполнитель	Оклад, руб./мес.	Среднедневная ставка, руб./раб.день	Затраты времени, раб. дни	Коэффициент	Фонд з/платы, руб.
НР	23264,86	933,58	14	1,699	22206,13
И	7864,11	315,57	94	1,62	48055,00
Итого					70261,13

9.2.3 Расчет затрат на социальный налог

К данной статье затрат относят единый социальный налог (ЕСН), который включают в себя отчисления в пенсионный фонд, на социальное и медицинское страхование. Отчисления по ЕСН определяются по следующей формуле:

$$C_{соц} = K_{соц} \cdot C_{осн} ,$$

где $K_{соц}$ – коэффициент отчислений. Значение данного коэффициента составляет 30 % от полной заработной платы по проекту.

$$C_{соц} = 0,3 \cdot 70261,13 = 21278,34.$$

9.2.4 Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную в ходе выполнения проекта на работу используемого оборудования. Затраты на электроэнергию рассчитываются по формуле:

$$C_{эл.об.} = P_{об} \cdot t_{об} \cdot ЦЭ,$$

где $P_{об}$ – мощность, потребляемая оборудованием, кВт;

$ЦЭ$ – тариф на 1 кВт·час;

$t_{об}$ – время работы оборудования, час.

Для ТПУ $ЦЭ = 5,782$ руб./кВт·час (с НДС).

Время работы оборудования вычисляется на основе итоговых данных таблицы 10 для инженера ($T_{РД}$) из расчета, что продолжительность рабочего дня равна 8 часов.

$$t_{об} = T_{РД} * K_t,$$

где $K_t \leq 1$ – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к $T_{РД}$. В данном случае значение коэффициента принимается равным 0,9.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{об} = P_{ном.} \cdot K_C$$

где $P_{ном.}$ – номинальная мощность оборудования, кВт;

$K_C \leq 1$ – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности данный коэффициент принимается равным 1.

Расчеты затрат на электроэнергию для технологических целей приведены в таблице 15.

Таблица 15 – Затраты на электроэнергию технологическую

Наименование оборудования	Время работы оборудования $t_{об}$, час	Потребляемая мощность $P_{об}$, кВт	Затраты $Э_{об}$, руб.
Персональный компьютер	$750,72 \cdot 0,9$	0,15	585,99
Струйный принтер	28	0,1	16,19
Итого:			602,18

9.2.5 Расчет амортизационных расходов

В данной статье расходов рассчитывается амортизация используемого оборудования за время выполнения проекта. Для этого используется формула:

$$C_{AM} = \frac{H_A \cdot C_{OB} \cdot t_{рф} \cdot n}{F_D},$$

где H_A – годовая норма амортизации единицы оборудования;

C_{OB} – балансовая стоимость единицы оборудования с учетом ТЗР;

F_D – действительный годовой фонд времени работы соответствующего оборудования;

$t_{рф}$ – фактическое время работы оборудования в ходе выполнения проекта, учитывается исполнителем проекта;

n – число задействованных однотипных единиц оборудования.

Так, H_A для персонального компьютера принимается равным 0,4, для струйного принтера $H_A = 0,5$.

Балансовая стоимость принимается равной действующей цене единицы оборудования, находящейся в прейскурантах. Для персонального компьютера $C_{OB} = 35000$ рублей, для струйного принтера $C_{OB} = 7800$ рублей.

Действительный годовой фонд времени работы оборудования берется из фактического режима его использования в текущем календарном году. Так, для персонального компьютера при 299 рабочих днях (шестидневная рабочая неделя) можно принять $F_D = 298 \cdot 8 = 2392$ часа. Для струйного принтера $F_D = 500$ часов.

$$C_{AM\text{ ПК}} = \frac{0,4 \cdot 35000 \cdot 750,72}{2392} = 4393,85 ;$$

$$C_{AM\text{ Пр}} = \frac{0,5 \cdot 7800 \cdot 28}{500} = 218,4 .$$

Итого начислено амортизации 4612,25 рублей.

9.2.6 Расчет расходов, учитываемых непосредственно на основе платежных (расчетных) документов (кроме суточных)

К данной статье расходов относятся командировочные расходы, арендная плата за использование имуществом, оплата услуг завязи и услуг сторонних организаций.

Расходы по данному пункту составляют $C_{нр} = 0$ рублей.

9.2.7 Расчет прочих расходов

В статье «Прочие расходы» отображены расходы на выполнение проекта, которые не учтены в предыдущих статьях. Данные расходы следует принять равными 10 % от суммы всех предыдущих расходов:

$$C_{\text{проч.}} = (C_{\text{мат}} + C_{\text{зн}} + C_{\text{соц}} + C_{\text{эл.об.}} + C_{\text{ам}} + C_{\text{нр}}) \cdot 0,1;$$

$$C_{\text{проч.}} = (1908,9 + 70261,13 + 21278,34 + 602,18 + 4612,25 + 0) \cdot 0,1 = 9915,49.$$

9.2.8 Цена разработки ВКР

Проведя расчет по всем статьям сметы затрат на разработку, можно определить общую себестоимость проекта. Смета на разработку проекта представлена в таблице 16.

Таблица 16 – Затраты на разработку проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	$C_{\text{мат}}$	1908,9
Основная заработная плата	$C_{\text{зн}}$	70261,13
Отчисления в социальные фонды	$C_{\text{соц}}$	21278,34
Расходы на электроэнергию	$C_{\text{эл.}}$	602,18
Амортизационные отчисления	$C_{\text{ам}}$	4612,25
Непосредственно учитываемые расходы	$C_{\text{нр}}$	0
Прочие расходы	$C_{\text{проч}}$	9860,81
Итого:		108523,61

9.2.9 Прибыль

Прибыль от реализации проекта составляет 20 % от расходов на разработку, т.е. $108523,61 \cdot 0,2 = 21704,72$ рублей.

9.2.10 Расчет НДС

НДС составляет 18% от суммы затрат на разработку и прибыли. В данном случае это $(108523,61 + 21704,72) \cdot 0,18 = 23441,09$ рублей.

9.2.11 Цена разработки ВКР

Цена разработки равна сумме полной себестоимости, прибыли и НДС. В данном случае $C_{\text{ВКР(КР)}} = 108523,61 + 21704,72 + 23441,09 = 153669,42$ рублей.

9.3 Оценка экономической эффективности проекта

Статистический анализ данных, использующийся для определения уровня равнозначности контрольно-измерительных материалов по дискретной математике, позволяет распределить на экзамене между студентами вопросы одинаковой сложности, чтобы оценка за экзамен зависела от знания студентом материала, а не от сложности вопроса

Разрабатываемая статистическая модель позволяет разделить вопросы на 3 уровня сложности, сравнить уровни сложности изучаемых тем в дискретной математике, сравнить результаты экзамена за 2015 и 2016 года и предсказать полученный студентом балл за экзамен по одному вопросу.

Оценка экономической эффективности проекта является некорректной, поскольку данная разработка не ориентирована на получение экономического эффекта.

9.4 Оценка научно-технического уровня ВКР

Научно-технический уровень характеризует влияние проекта на уровень и динамику обеспечения научно-технического прогресса в данной области. Для оценки научной ценности, технической значимости и эффективности, планируемых и выполняемых ВКР воспользуемся методом бальных оценок:

$$K_{НТУ} = \sum_{i=1}^3 R_i \cdot n_i ,$$

где $I_{НТУ}$ – интегральный индекс научно-технического уровня;

R_i – весовой коэффициент i -го признака научно-технического эффекта;

n_i – количественная оценка i -го признака научно-технического эффекта, в баллах.

Таблица 17 – Оценки научно-технического уровня ВКР

Фактор НТУ	Значимость	Уровень фактора	Выбранный балл	Обоснование выбранного балла
Уровень новизны	0,4	Новая	6	Аппарат математической статистики применяется к

				контрольно-измерительным материалам по дискретной математике
Теоретический уровень	0,1	Глубокая разработка проблемы, многоспектральный анализ взаимодействия между факторами с наличием объяснений	8	Данная статистическая модель определяет уровень равнозначности контрольно-измерительных материалов по дискретной математике.
Возможность реализации	0,5	В течение первых лет	10	Полная реализация в течение одного года

Интегральный показатель научно-технического уровня для данного проекта составляет:

$$0,4 \cdot 6 + 0,1 \cdot 8 + 0,5 \cdot 10 = 2,4 + 0,8 + 5 = 8,2$$

Исходя из полученного значения можно сделать вывод, что данный проект имеет высокий уровень научно-технического эффекта.

СПИСОК ПУБЛИКАЦИЙ СТУДЕНТА

1. Чернета Д. С. Применение кластерного и дисперсионного анализа для оценки качества обучения по дискретной математике // Молодежь и современные информационные технологии: сборник трудов XIV Международной научно-практической конференции студентов, аспирантов и молодых ученых: в 2 т., Томск, 7-11 ноября 2016. – Томск: ТПУ, 2016 - Т. 1 – С. 212-213