1 Цели и задачи учебного модуля

Целью учебного модуля (УМ): «Интеллектуальный анализ данных и введение в теорию некорректных задач» является

формирование представления о типах задач, возникающих в области интеллектуального анализа данных (*DataMining*) и методах их решения, которые помогут студентам выявлять, формализовать и успешно решать практические задачи анализа данных, возникающие в процессе их профессиональной деятельности;

ознакомление обучаемых с проблемой корректности и обусловленности вычислительных задач, освоение студентами методик определения корректности и обусловленности вычислительных задач, освоение студентами методик выбора параметра регуляризации и примерах СЛАУ. Курс призван дать обзор некоторых актуальных научных проблем прикладной математики и информатики, а также существующих в настоящее время методов, подходов и средств решения данных проблем.

Для достижения цели перед магистрантами ставятся следующие **задачи**:

- освоение теоретического базиса дисциплины;
- воспитание у студента умения применять полученные знания при исследовании практических задач, культуры мышления;
- развитие у студента математической культуры и интуиции;
- привитие студентам навыков самостоятельной работы по изучению специальной литературы;
- формирование математической культуры самостоятельного применения полученных знаний в реальных исследованиях;
- формирование представления о типах задач, возникающих в области интеллектуального анализа данных (*DataMining*) и методах их решения, которые помогут студентам выявлять, формализовать и успешно решать практические задачи анализа данных, возникающие в процессе их профессиональной деятельности.
- воспитание у студента умения формулирования и обоснования выбора математической модели;
- ознакомление студентов с проблемами, требующими математического моделирования;
- формирование у студентов практических навыков разработки математических моделей;
- формирование математической культуры самостоятельного применения полученных знаний в реальных исследованиях

Решение указанных задач проводится с использованием примеров, взятых из различных областей.

2 Место учебного модуля в структуре ОП направления подготовки

Модуль «Интеллектуальный анализ данных и введение в теорию некорректных задач» входит в учебный план подготовки магистров по направлению 01.04.02 - "Прикладная математика и информатика" в базовую часть, трудоемкость 6 зачетных единицы.

Содержание модуля тесно связано со всеми дисциплинами, которые изучались студентами. Предполагается, что студенты владеют основными понятиями математического анализа, дискретной математики, линейной алгебры, теории групп, численных методов, теории вероятностей и математической статистики и др. Т.е. изучение курса «Интеллектуальный анализ данных и введение в теорию некорректных задач» требует предварительных знаний в объеме стандартной программы бакалавриата.

В общеобразовательной программе магистранта данная дисциплина является необходимым средством выполнения выпускной научной работы магистранта, написания выпускной квалификационной работы.

3 Требования к результатам освоения модуля

Изучение дисциплины направлено на формирование перечисленных ниже элементов общекультурных и профессиональных компетенций:

- готовностью действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения (ОК-2);

– готовность к саморазвитию, самореализации, использованию творческого потенциала (ОК-3);

– способность разрабатывать и применять математические методы, системное и прикладное программное обеспечение для решения задач научной и проектно-технологической деятельности (ПК-3);

В результате освоения УМ студент должен знать, уметь и владеть:

Код компетенции	Уровень освоения компетенции	Знать	Уметь	Владеть
ОК2	Базовый	определение понятий социальной и этической ответственности при принятии решений, различие форм и последовательности действий в стандартных и некоторых нестандартных ситуациях	: анализировать варианты действий в некоторых нестандартных ситуациях, определять меру социальной и этической ответственности за принятые решения	: целостной системой навыков действий в некоторых нестандартных ситуациях, прогнозировать результаты социальной и этической ответственности за принятые решения
ОК3	базовый	ЗНАТЬ: содержание процесса формирования основных целей профессионального и личностного развития, некоторые способы его реализации при решении профессиональных задач, подходы и существенные ограничения при использовании творческого потенциала	формулировать основные цели личностного и профессионального развития и условия их самореализации с учётом индивидуально-личностных особенностей и возможностей использования творческого потенциала	основными приемами и технологиями формирования целей саморазвития и их самореализации, критической оценки результатов деятельности по решению профессиональных задач и использованию творческого потенциала
ПК3	повышенный	основные моменты в современном состоянии и проблемах в области интеллектуального анализа данных (<i>DataMining</i>), в области некорректных задач, о основных типах задач, возникающих в области интеллектуального анализа данных (<i>DataMining</i>), в области некорректных задач и методах их решения	Разрабатывать обзоры о современном состоянии и проблемах в области интеллектуального анализа данных (<i>DataMining</i>), в области некорректных задач, самостоятельно приобретать новые знания в области интеллектуального анализа данных (<i>DataMining</i>), в области некорректных задач; Разрабатывать основные модели в области интеллектуального анализа данных и в области некорректных задач	Основными моментами в области интеллектуального анализа данных и в области некорректных задач,

4 Структура и содержание дисциплины

4.1 Трудоемкость дисциплины и формы аттестации

Таблица 1.1 – Трудоемкость дисциплины в зачетных единицах и распределение трудоемкости по видам учебной работы и семестрам в академических часах. Дневная форма обучения для направления 01.04.02 – прикладная математика и информатика

Учебная работа (УР)	Всего	Распределение по семестрам
		3
Полная трудоемкость дисциплины в зачетных единицах (ЗЕ), в том числе:	6	6
Распределение трудоемкости по видам УР в академ. часах (АЧ):	-	-
– лекции	9	9
– практические занятия (семинары)	36	36
– в том числе, аудиторная СРС	9	9
– внеаудиторная СРС	171	171
Аттестация:– Экзамен		

4.2 Содержание и структура разделов учебного модуля

В структуре УМ выделены следующие учебные элементы модуля (УЭМ) в качестве самостоятельных разделов:

УЭМ1Интеллектуальный анализ данных 4/16/4/71

1. Введение в интеллектуальный анализ данных 1. Анализ, выбраковка и восстановление аномальных или пропущенных измерений. 1/4/1/15

2 Оценка параметров и числовых характеристик наблюдаемых случайных величин или процессов.

3Группировка исходной информации при большом объеме обрабатываемых данных.). 1/4/1/15

4 Выявление скрытых связей и взаимовлияния различных измеряемых факторов и результирующих переменных. 1/4/1/15

5 Выдвижение гипотез о классе и структуре математической модели исследуемого явления 0/0/0/11

УЭМ2Введение в теорию некорректных задач5/20/5/100

1. Элементы функционального анализа и теории алгебраических структур.Погрешности элементов нечисловых множеств1/4/1/20

2. 2 Корректность и обусловленность вычислительной задачи. Приложения теории некорректных задач. 1/4/1/20

3. Корректность и обусловленность . решения СЛАУ. Регуляризация решения СЛАУ. 2/8/2/40

4. Корректность и обусловленность вычислительного алгоритма 1/4/1/20

Календарный план, наименование разделов учебного модуля с указанием трудоемкости по видам учебной работы представлены в технологической карте учебного модуля (приложение Б).

4.3 Организация изучения учебного модуля

Методические рекомендации по организации изучения УМ с учетом использования в учебном процессе активных и интерактивных форм проведения учебных занятий даются в Приложении А

5 Контроль и оценка качества освоения учебного модуля

Контроль качества освоения студентами УМ «Интеллектуальный анализ данных и введение в теорию некорректных задач» и его составляющих осуществляется непрерывно в течение всего периода обучения с использованием балльно-рейтинговой системы (БРС), являющейся обязательной к использованию всеми структурными подразделениями университета.

Для оценки качества освоения модуля используются формы контроля: текущий – регулярно в течение всего семестра, рубежный и семестровый (экзамен) – по окончании изучения УМ.

Рубежный контроль проходит на 18 неделе, по окончании изучения УМ и осуществляется посредством экзамена и подсчетом суммарных баллов за весь период изучения УМ. Пороговому уровню соответствует 125 баллов, максимальное количество баллов за рубежный контроль – 250 баллов. Максимальное количество баллов, получаемое на экзамене, – 50. Максимальное количество баллов, полученных при изучении УМ – 300 баллов.

Оценка качества освоения модуля осуществляется с использованием фонда оценочных средств, разработанного для данного модуля, по всем формам контроля в соответствии с положением от 27.09.2011 № 32 «Об организации учебного процесса по основным образовательным программам высшего профессионального образования».

В качестве оценочных средств на протяжении семестра используются: разноуровневые задачи, опросы, индивидуальные домашние задания, контрольные работы, тесты.

Содержание видов контроля и их график отражены в технологической карте учебного модуля (Приложение Б). Паспорта компетенций представлены в приложении В.

6 Учебно-методическое и информационное обеспечение

Учебно-методическое и информационное обеспечение учебного модуля представлено Картой учебно-методического обеспечения (Приложение Г).

7 Материально-техническое обеспечение учебного модуля

Для осуществления образовательного процесса по модулю используется лекционная аудитория, оборудованная мультимедийными средствами, а также лаборатория (компьютерный класс), лицензионное программное обеспечение.

Программное обеспечение:

1. Математическая система *MathCAD 13.0 (MathSoft)* или Профессиональная среда для выполнения вычислений *Maple (WaterlooMapleSoftware)*.
2. Электронные таблицы *Excel*.
3. Системы программирования: *TurboPascal, BorlandC++*.
4. , *MicrosoftOffice*;
5. бесплатные версии специализированных программных средств анализа данных , *SPSS, "STATISTICADATAMINER", "PolyAnalist", "Deductor "*

Приложения (обязательные):

- А – Методические рекомендации по организации изучения учебного модуля
- Б – Технологическая карта
- В – Карта учебно-методического обеспечения УМ
- Г – Лист внесения изменений

Приложение А
Методические рекомендации по организации изучения учебного модуля
«Интеллектуальный анализ данных и введение в теорию некорректных задач»
(обязательное)

Учебный модуль «Интеллектуальный анализ данных и введение в теорию некорректных задач» разделен на два учебных элемента модуля (УЭМ):

УЭМ1 Интеллектуальный анализ данных;

УЭМ2 Введение в теорию некорректных задач.

Каждый из УЭМ состоит из взаимосвязанных разделов, по которым предусмотрены лекционные и практические занятия. В таблице А.1 отражены разделы модуля, технологии и формы проведения занятий, задания по самостоятельной работе студента и ссылки на необходимую литературу.

А.1 Методические рекомендации по теоретической части учебного модуля

Теоретическая часть модуля направлена на формирование системы знаний об основных разделах современной прикладной математики и информатики с выделением аспектов применения современного математического аппарата в исследовательской деятельности. Основное содержание теоретической части излагается преподавателем на лекционных занятиях, а также усваивается студентом при знакомстве с дополнительной литературой, которая предназначена для более глубокого овладения знаниями основных дидактических единиц соответствующего раздела и указана в таблице А.1.

А.2 Методические рекомендации по практическим занятиям

Цель практических занятий – закрепление теоретического материала и выработка у студентов умения решать задачи по практическим аспектам учебного модуля.

На практических занятиях студентам предлагаются задачи и вопросы по пройденному разделу дисциплины. На занятиях преподаватель проверяет выполнение домашних заданий, разбирает вместе со всеми нерешенные дома задачи.

Практические занятия в рамках УЭМ1-2 строятся следующим образом:

- 20% аудиторного времени отводится на объяснение решения типовой задачи у доски;
- 70% аудиторного времени – самостоятельное решение задач студентами;
- 10% аудиторного времени в конце текущего занятия – разбор типовых ошибок при решении задач.

При подготовке к практическим занятиям студент должен изучить лекционный материал, в случае необходимости обратиться к соответствующим разделам рекомендованной литературы и методическим пособиям, разработанным на кафедре ПМИ. При изучении материала необходимо отметить вызывающие затруднения вопросы для получения консультации у преподавателя. К практическим занятиям по конкретной теме студент обязан знать основные понятия, определения, формулировки теорем и свойства. На практических занятиях необходимо иметь конспект лекций по изучаемой теме.

А.3 Методические рекомендации по самостоятельной работе студентов

Для подготовки к практическим занятиям, контрольной работе, и экзамену рекомендуется пользоваться основной и дополнительной учебно-методической литературой, представленной в карте учебно-методического обеспечения. Для закрепления темы студенту выдаются индивидуальные домашние задания (ИДЗ) для самостоятельной работы. Они выполняются на отдельных листах и защищаются во время аудиторной СРС. ИДЗ выдаётся на первом практическом занятии по разделу и выполняется по мере изучения материала. При их выполнении рекомендуется использовать проработанный в аудитории материал и обратиться к задачкам, в которых разобраны типовые примеры с решениями стандартных задач. Таким образом, после

каждого практического занятия студент закрепляет пройденный материал.

Таблица А. Организация изучения каждого УЭМ для УМ «Интеллектуальный анализ данных и введение в теорию некорректных задач»

УЭМ	Технология и форма проведения занятий, СРС	Задания на СРС	Литература
УЭМ 1, 2 ¹	<ul style="list-style-type: none"> - вводная лекция, - информационная лекция, - обзорно-проблемная лекция, - решение типовых заданий преподавателем, магистрантами под руководством преподавателя, выполнение индивидуальных заданий, - разбор заданий индивидуальной работы, анализ ее результатов. 	<ul style="list-style-type: none"> -используя основную и дополнительную литературу, изучить основные дидактические единицы разделов (ВСРС), - проанализировать типовые модели при изменении данных (АСРС и ВСРС), - выполнить задания по теме УЭМ (ВСРС) 	<ul style="list-style-type: none"> 1 Дюк В.А., Самойленко А.П. DataMining: учебный курс. — СПб. : Питер, 2001. - 366 с Чубукова И.А. DataMining : учеб.пособие / И. А. Чубукова. - 2-е изд., испр. - М. : Интернет-университет Информ. Технологий : БИНОМ. Лаборатория знаний, 2008. - 382 с Ольховой А.Ф. 'Обратные некорректные задачи. Введение в проблематику' - Таганрог: Технологический институт ЮФУ ГСП 17А, 2009 - с.131

Вопросы по дисциплине для самостоятельной подготовки к экзамену

УЭМ1

1. . Понятие интеллектуального анализа данных. Методы DataMining.
2. Анализ, выбраковка и восстановление аномальных или пропущенных измерений.
3. Понятие разведочного анализа данных. Отличие процедуры DataMining от методов классического статистического анализа данных?
4. Методы графического разведочного анализа данных. Средства Statistica для проведения графического разведочного анализа данных.
5. Достоинства и недостатки графических методов разведочного анализа данных
6. Аналитические методы первичного разведочного анализа данных
7. Проверка гипотез о согласии распределения выборочных данных с моделью нормального распределения?
8. Основные статистические характеристики количественных переменных и их описание и интерпретация в терминах решаемой задачи.
9. Коэффициент эксцесса (близость к нормальному распределению)
10. Измерители связи применяются для измерения степени тесноты связи между количественными и порядковыми переменными? Их расчет в Statistica и интерпретация

УЭМ2

1. Корректность вычислительной задачи. Примеры корректных и некорректных задач
2. Корректность вычислительной задачи суммирования сходящегося ряда в случае различного определения погрешностей входных данных
3. Некорректность решения уравнения Фредгольма 1-го рода.
4. Обусловленность вычислительной задачи. Примеры хорошо и плохо обусловленных задач.
5. Обусловленность задачи нахождения кратных корней
6. Обусловленность задачи нахождения корней многочлена высокой степени

¹ Содержание и структура разделов представлены в п. 4.2 РП для УМ.

7. Число обусловленности матриц. Свойства числа обусловленности матрицы
8. Обусловленность задачи решения СЛАУ при изменении правой части.
9. Обусловленность задачи решения СЛАУ при изменении матрицы данных
10. Обусловленность задачи решения СЛАУ при изменении матрицы данных и правой части.
11. МНК как решение некорректной задачи
12. Определение регуляризирующего функционала А.Н. Тихонова
13. Некорректные СЛАУ.
14. Определение нормального решения вырожденных СЛАУ.
15. Метод регуляризации СЛАУ.
16. Корректность и обусловленность вычислительных алгоритмов. Корректность алгоритма вычисления по рекуррентным формулам
17. Корректность и обусловленность вычислительных алгоритмов. Корректность алгоритма вычисления частичной суммы ряда при отсутствии ошибок компьютерного округления данных.
18. Корректность и обусловленность вычислительных алгоритмов. Корректность алгоритма вычисления частичной суммы ряда при наличии ошибок компьютерного округления данных.

Вопросы для подготовки к самостоятельной работе

1. Понятие интеллектуального анализа данных. Методы DataMining.
2. Понятие разведочного анализа данных. В чем отличие процедуры DataMining от методов классического статистического анализа данных?
3. Методы графического разведочного анализа данных. Средства Statistica для проведения графического разведочного анализа данных.
4. Что такое категоризованный график? Какие типы категоризованных графиков реализованы в системе Statistica?
5. Какую информацию о природе данных можно получить при анализе диаграмм рассеяния и категоризованных диаграмм рассеяния?
6. Какую информацию о природе данных можно получить на основе анализа гистограмм и категоризованных гистограмм?
7. Чем принципиально отличаются категоризованные графики от матричных графиков?
8. В чем достоинства и недостатки графических методов разведочного анализа данных?
9. Какие аналитические методы первичного разведочного анализа данных вы знаете?
10. Как проверить гипотезу о согласии распределения выборочных данных с моделью нормального распределения?
11. Какие основные статистические характеристики количественных переменных вы знаете? Их описание и интерпретация в терминах решаемой задачи. коэффициента эксцесса (близость к нормальному распределению)
12. Какие измерители связи применяются для измерения степени тесноты связи между количественными и порядковыми переменными? Их расчет в Statistica и интерпретация.

Индивидуальные задания

ЗАДАНИЕ 1

Обработка данных, в том числе и результатов эксперимента, является важнейшим средством получения новых знаний не только в области естественных и технических наук, но и в экономике, социологии, политике, психологии, литературоведении и в других отраслях. Эти исследования дают критерии оценки обоснованности и приемлемости на практике любых теорий и теоретических предположений. Обработка данных направлена,

как правило, на построение математической модели исследуемого объекта или явления, а также на получение ответа на вопрос: «Достоверны ли имеющиеся данные в пределах требуемой точности или допусков?».

Сама же математическая модель в зависимости от целей (исследование, управление, контроль) может быть использована для разных целей: для предметно-смыслового анализа объекта или явления, прогнозирования их состояния в разных условиях функционирования, управления ими в конкретных ситуациях, оптимизации отдельных параметров, а также для решения каких-то других специфичных задач.

Конечной целью любой обработки данных является выдвижение гипотез о классе и структуре математической модели исследуемого явления, определение состава и объема дополнительных измерений, выбор возможных методов последующей статистической обработки и анализ выполнения основных предпосылок, лежащих в их основе. Для ее достижения необходимо решить некоторые частные задачи, среди которых можно выделить следующие:

1. Анализ, выбраковка и восстановление аномальных или пропущенных измерений.

Эта задача связана с тем, что исходная информация обычно неоднородна по качеству. В основной массе результатов прямых измерений, получаемых с возможно малыми погрешностями, в имеющихся данных часто имеются грубые ошибки или просчеты, вызванные разными причинами. К ним могут быть отнесены особенности информационной системы, сбои вычислительной техники, аномалии в работе измерительных приборов и т. д. Без глубокого анализа качества данных, устранения или, хотя бы, существенного уменьшения влияния аномальных данных на результаты последующей обработки можно сделать ложные выводы об изучаемом объекте или явлении.

2. Оценка параметров и числовых характеристик наблюдаемых случайных величин или процессов.

Выбор методов последующей обработки, направленной на построение и проверку адекватности математической модели исследуемому явлению, существенно зависит от закона распределения наблюдаемых величин. Получаемые при решении этой задачи выводы о природе обрабатываемых данных могут быть как общими (независимость измерений, характер погрешностей и др.), так и содержать детальную информацию о свойствах данных (в том числе и статистических, таких как вид закона распределения, его параметры). Решение задачи предварительной обработки не является чисто математическим, а требует также и содержательного анализа изучаемого процесса, а при возможности, схемы и методики проведения эксперимента.

3. Группировка исходной информации при большом объеме обрабатываемых данных.

Должны быть учтены особенности их законов распределения, которые выявлены на предыдущем этапе .

4. Выявление скрытых связей и взаимовлияния различных измеряемых факторов и результирующих переменных.

Решение этой задачи позволяет отобрать те переменные, которые оказывают наиболее сильное влияние на результирующий признак. Выделенные факторы используются для дальнейшей обработки, в частности, методами регрессионного анализа,

МГК. Анализ корреляционных связей делает возможным выдвижение гипотез о структуре взаимосвязи переменных и, в конечном итоге, о структуре модели объекта исследований.

В ходе предварительной обработки, часто решают и другие, имеющие частный характер задачи: преобразование и унификацию типа наблюдений, визуализацию многомерных данных и др.

1. СОЗДАНИЕ БАЗЫ ДАННЫХ.

Из сборников Росстата «Здравоохранение» составить базу данных указанных переменных для 2001, 2005, 2007, 2009, 2011, 2013, 2015 гг (в Экселе). Всего рассматриваем 7 наблюдений (для семи лет).

Переменные должны быть описаны таким образом, чтобы сторонний исследователь мог получить те же цифры и проверить самостоятельно приводимые рассуждения (см. пример ниже). В этом же описании должны содержаться комментарии о возможных исправлениях данных при устранении пропусков, корректировке выбросов и т.д.

Список субъектов составляется исходя из состояния на 2015 год. В справочниках за 2007 год и ранее имеются субъекты, которые далее упразднены.

Например, **Усть-Ордынский Бурятский автономный округ** упразднён 1 января 2008 года. Округ был объединён с Иркутской областью. Следовательно, в итоговом списке субъектов этот округ не значится, а данные для Иркутской области для 2001, 2005, 2007 гг пересчитываются с учетом численности населения для этих субъектов на момент наблюдения.

Крым в наши наблюдения не входит, т.к. по нему нет достаточного количества данных.

Если переменные даны в абсолютных величинах (количество в штуках больных), приводим их к относительной величине – на 1000 (может, на 10 000 или пр.) жителей, используя, опять же, данные Росстата для соответствующего года и субъекта Федерации.

2. УСТРАНЕНИЕ ВЫБРОСОВ И ПРОПУЩЕННЫХ ЗНАЧЕНИЙ

Для **каждой переменной** установить наличие/отсутствие выбросов (по горизонтали, сравнивая наблюдения для различных лет для каждого субъекта РФ):

- а) графически
- б) через значение межквартильного диапазона
- с) с помощью критериев Греббса (Смирнова), Шарлье, Шовенэ, Диксона (и еще наверное есть.....)

Там, где наличие выброса можно считать достоверным, исправить исходные данные. Также следует устранить возможные пропуски данных на основе наблюдения для этого субъекта по 7 годам и на основе значения наблюдаемого параметра для того Федерального округа, к которому относится субъект.

После выполнения этого шага должны быть 2 таблицы: с исходными данными с пропусками и выбросами и исправленные данные. Исправления в таблицах должны быть хорошо заметны, их выделяем цветом, шрифтом и т.д.

3. ОПИСАТЕЛЬНАЯ СТАТИСТИКА И ПРЕДСТАВЛЕНИЕ ДАННЫХ

Изобразить данные возможными графическими способами (красиво и наглядно, используя ВСЕ ВОЗМОЖНОСТИ ПАКЕТА...).

С помощью статистического пакета вычислить все возможные описательные статистики, обращая внимание стандартные ошибки вычисляемых параметров и на доверительные интервалы для них.

Пример некоторых показателей, вычисляемых пакетом Statistika.

- *Valid N* – объем выборки;
 - *Mean* – арифметическая средняя;
 - *Sum* – сумма значений анализируемой переменной;
 - *Median* – медиана;
 - *Mode* – мода;
 - *Geom. mean* – геометрическая средняя;
 - *Harm. mean* – гармоническая средняя;
 - *Standard Deviation* – стандартное отклонение;
 - *Variance* – дисперсия;
-
- *Std. err. of mean* – стандартная ошибка средней;
 - *Conf. limits for means: Interval %* – доверительные пределы для средних: ширина доверительного интервала;
 - *Skewness* – коэффициент асимметрии;
 - *Std. err., Skewness* – стандартная ошибка коэффициента асимметрии;
 - *Kurtosis* – коэффициент эксцесса;
 - *Std. err., Kurtosis* – стандартная ошибка коэффициента эксцесса;
 - *Minimum & maximum* – минимальное и максимальное значения;

Таким образом, статистический пакет дает гораздо больше инструментов, чем Эксель. Сделать выводы по полученным значениям.

ЗАДАНИЕ2

4 ПРОВЕРКА ДАННЫХ НА НОРМАЛЬНОСТЬ

ЛЮБОЕ изучение количественного параметра начинается с изучения его распределения. Одним из важных моментов является проверка распределения на соответствие предположения о его нормальности. Исследуем распределение величин – для всей совокупности (по семи наблюдениям) и для выборок по каждому году.

Способы:

- Графический способ (изучение гистограммы распределения, построение квантильных диаграмм).
- Численные способы (изучение характеристик распределения: среднее арифметическое, гармоническое, медиана; дисперсия, коэффициенты вариации, ангармонизма, эксцесса)
- Проверки гипотезы о нормальности распределения (количественные тесты):
 - Колмогорова-Смирнова, Лиллифорс, Шапиро-Уилка (стандартно);
 - более современные: Шапиро-Франка, Андерсона-Дарлинга, Крамера-фон-Мизеса.

Кроме формальной проверки распределения на нормальность с имеющимися данными, выполнить нормализацию данных обратимыми функциями:
логарифмирование,
преобразование Бокса-Кокса и др.
с последующим изучением распределения преобразованных данных.

Следует предостеречь от формального «соответствует/не соответствует». Как правило, мы не знаем истинного распределения данных. На основе выполненного анализа исследователь принимает решение о соответствии или несоответствии распределения. Т.е., ВСЕГДА речь идет о принятом решении.

С помощью STATISTICA можно быстро оценить, достаточно ли хорошо реальное распределение моделируется некоторым теоретическим. При этом алгоритм такой оценки остаётся за кадром. Это иногда может быть недостатком. В этом случае полезно использовать *MS Excel*, что позволяет выполнять оценку пошагово, усваивая сущность алгоритма, но избавляясь от рутинных расчётов. При этом можно создавать пересчитываемые электронные таблицы, автоматически выдающие результат при вводе новых данных. Методики оценки распределения при помощи *MS Excel* несколькими способами можно найти на <http://arhiuch.ru>.

5 ИССЛЕДОВАНИЕ РАЗЛИЧИЙ В ВЫБОРКАХ

Проверить, можно ли объяснить имеющееся различие средних значений статистическими колебаниями или нет. Т.е. проверить наличие о **статистически значимого различия** между сравниваемыми группами.

- a) Между федеральными округами для каждого наблюдения (по вертикали)
- b) До и после кризиса 2008 года. (по горизонтали)

При сравнении **средних значений** выборок предполагается, что обе выборки **подчиняются нормальному распределению**. Если это не так, то вычисляются **медианы** и для сравнения выборок используется **непараметрические тесты**.

6 ПОСТРОЕНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Исследуем возможность зависимости переменных методами регрессионного и многофакторного анализа

В качестве зависимой переменной рассматриваем Смертность населения по субъектам Российской Федерации и Заболеваемость населения по субъектам Российской Федерации, в качестве независимых – рассматриваемые переменные.

1. Сайт Росстата http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1139919134734
2. <http://vnauke.by/schkola/Osnovy-dokazatelnoj-mediciny-Biomedicinskaja-statistika/Proverka-normalnosti-raspredeleniya-znachenij-priznaka>
3. http://cdn.scipeople.com/materials/7765/%D0%9C%D0%B0%D1%81%D1%82%D0%B8%D1%86%D0%BA%D0%B8%D0%B9_2009_STATISTICA6.pdf
4. Статистический пакет STATISTIKA <https://mirznanii.com/a/16472/statisticheskiy-paket-statistika>

Таблица 3 – Переменные для выполнения задания

№ варианта	Зависимые переменные	Независимые переменные	№ Таблицы в справочнике за 2015 г
1	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам РФ	Заболеваемость населения по основным классам болезней по субъектам РФ	Т.2.2
2	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам	. Заболеваемость населения отдельными инфекционными болезнями по субъектам РФ	Т. 2.3
3	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам	Заболеваемость населения социально-значимыми болезнями по субъектам РФ	Т. 2.4
4	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам	Основные экономические показатели здравоохранения	Т. 7.1- 7.11
5	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам	Контингенты пациентов с болезнями....	Т. 2.5 – 2.14
6	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам РФ	Заболеваемость населения по основным классам болезней по субъектам РФ	Т.2.2
7	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам	. Заболеваемость населения отдельными инфекционными болезнями по субъектам РФ	Т. 2.3
8	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам	Заболеваемость населения социально-значимыми болезнями по субъектам РФ	Т. 2.4
9	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам	Основные экономические показатели здравоохранения	Т. 7.1- 7.11
10	Смертность населения по субъектам РФ, всего, Заболеваемость населения по субъектам	Контингенты пациентов с болезнями....	Т. 2.5 – 2.14

Министерство науки и высшего образования Российской Федерации
 Новгородский государственный университет имени Ярослава Мудрого
 Институт электронных и информационных систем
 Кафедра прикладной математики и информатики
ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №2
 по курсу «ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ
 И ВВЕДЕНИЕ В ТЕОРИЮ НЕКОРРЕКТНЫХ ЗАДАЧ

1. Корректность вычислительной задачи. Корректность вычислительной задачи суммирования сходящегося ряда в случае различного определения погрешностей входных данных

2. Оценить обусловленность системы , заданной матрицами

$$A^* = \begin{pmatrix} 1 & 1 \\ 1 + \mu & 1 \end{pmatrix}, B^* = \begin{pmatrix} 1 \\ 1 + \delta \end{pmatrix}$$

3 Решить систему методом регуляризации

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Зав. кафедрой ПМИ

А.В. Колногоров

Приложение Б
Технологическая карта

УМ «Интеллектуальный анализ данных и введение в теорию некорректных задач»

семестр 3, ЗЕТ 6, вид аттестации: экзамен

Акад. часов 216 , баллов рейтинга 300

№ и наименование раздела учебного модуля, КП/КР	№ неде- ли сем.	Трудоемкость, ак.час					С Р С	Форма текущего контроля успев. (в соотв. с паспорт м ФОС)	Макси м.кол- во баллов рейти нга
		Аудиторные занятия				А СРС			
		ЛЕК	ПЗ	ЛР					
УЭМ1Интеллектуальный анализ данных	1-9	4	16			4	71	ИЗ1 СРС1 СРС2 СРС3	65 20 20 20
Рубежный контроль		Рубежный контроль – не менее 63 баллов из 125							
УЭМ2Введение в теорию некорректных задач	10-14	5	20			5	100	ИЗ2 СРС3 СРС4 СРС5	65 20 20 20
Семестровый контроль Экзамен		50							
Итого:		9	36	0		9	171		300

Критерии оценки качества освоения студентами дисциплины:

оценка «удовлетворительно») – 150– 207 баллов;

оценка «хорошо») – 208 - 267баллов;

оценка «отлично») – 268 – 300 баллов.

Приложение В Карта учебно-методического обеспечения

Дисциплины «Интеллектуальный анализ данных и введение в теорию некорректных задач»

для направления **01.04.02** – прикладная математика и информатика

Форма обучения очная Курс 1 Семестр 2

Всего часов 216 из них: лекций 9 часов, практические занятия – 36 часов, СРС – 171 час.,

Обеспечивающая кафедра ПМИ Институт *ИЭИС*

Таблица 1- Обеспечение учебного модуля учебными изданиями

Библиографическое описание* издания (автор, наименование, вид, место и год издания, кол.стр.)	Кол.экз. в библ. НовГУ	Наличие в ЭБС
Учебники и учебные пособия		
Дюк В.А., Самойленко А.П. DataMining: учебный курс. — СПб. : Питер, 2001. - 366 с.)	15	-
Чубукова И.А. DataMining : учеб.пособие / И. А. Чубукова. – 2-е изд., испр. – М. : Интернет-университет Информ. Технологий : БИНОМ. Лаборатория знаний, 2008. – 382 с.	12	Библиотех IPRbooks, ЭБС Лань
Ольховой А.Ф. Обратные некорректные задачи. Введение в проблематику [Электронный ресурс] – Таганрог: Технологический институт ЮФУ ГСП 17А, 2009 - с.131. – Режим доступа: http://mathemlib.ru/books/item/f00/s00/z0000013/index.shtml	-	ЭБС mathemlib.r u
Учебно-методические издания		
1 Рабочая программа модуля с приложениями «Интеллектуальный анализ данных и введение в теорию некорректных задач» » /Авт.-сост. Т.В. Жгун ; НовГУ. – В.Новгород, 2018. – 17 с.	-	есть электр. вариант

Таблица 2 – Информационное обеспечение учебного модуля

Название программного продукта, интернет-ресурса	Электронный адрес	Примечание
Электронная библиотека механико-математического факультета Московского государственного университета [Электронный ресурс] : [офиц. сайт] / Моск. гос. ун-т. – М., 2002-2015. – Режим доступа: http://www.lib.mechmat.ru/ , свободный. - Загл. с экрана (дата обращения: 17.02.2017).	http://poiskknig.ru	электронная библиотека учебников Мех-Мата МГУ, Москва
Общероссийский математический портал Math-Net.Ru [Электронный ресурс] / Математич. институт им. В. А. Стеклова РАН. - М.; 2002-2015. - Режим доступа: http://www.mathnet.ru/about.phtml?option_lang=rus , свободный - Загл. с экрана (дата обращения: 17.02.2017).	http://www.mathnet.ru/	общероссийский математический портал
WileyInterScience - Журналы издательства JohnWiley&Sons [Электронный ресурс] : [офиц. сайт] / JohnWiley&Sons, Ltd – Великобритания, 2002-2015. - Режим доступа: http://www3.interscience.wiley.com/ , свободный. - Загл. с экрана (дата обращения: 17.02.2017).	http://onlinelibrary.wiley.com	Журналы издательства JohnWiley&Sons
ScienceDirect [Электронный ресурс] : [офиц. сайт] / Издательство Elsevier (Эльзевир)– Великобритания, 2002-2014. - Режим доступа:	http://www.sciencedirect.com/	научные журналы издательства Elsevier

http://www.sciencedirect.com/ , свободный. - Загл. с экрана (дата обращения: 17.02.2017).		
Нечеткая логика, мягкие вычисления и вычислительный интеллект [Электронный ресурс] : [официальный сайт] / И. Батыршин, Российская Ассоциация Нечетких Систем. – М., 2004-2005. - Режим доступа: http://fuzzyset.narod.ru/Bisc.html /, свободный. - Загл. с экрана (дата обращения: 17.02.2017).	// http://fuzzyset.narod.ru	Сайт Российской Ассоциации Нечетких Систем и Мягких Вычислений (РАНСМВ)
InfTech. InformationTechnologies. Информационные Технологии [Электронный ресурс] : [официальный сайт] / Режим доступа: http://www.inftech.webservis.ru/ /, свободный. - Загл. с экрана (дата обращения: 17.02.2017).	http://inftech.webservis.ru/it/conference/isanditc/2000/session3/rus/arrus16.html	Сайт информационных технологий
VisualStudio[Электронный ресурс] : [официальный сайт] / Microsoft-2015. - Режим доступа: http://www.visualstudio.com/ru-ru/ , свободный. - Загл. с экрана (дата обращения: 17.02.2017).	http://www.microsoft.com/visualstudio/ru-ru	Среда программирования Microsoft Visual Studio 2010
Lazarus [Электронный ресурс] : [официальный сайт] LazarusandFreePascalTeam , -1993-2013. - Режим доступа: http://www.lazarus.freepascal.org/ свободный. - Загл. с экрана (дата обращения: 17.02.2017).	http://www.lazarus.freepascal.org/	Среда программирования Lazarus
Интернет-университет информационных технологий [Электронный ресурс] : [официальный сайт] / НОУ «ИНТУИТ», 2003 – 2015. - Режим доступа: http://www.intuit.ru/ свободный. - Загл. с экрана (дата обращения: 17.02.2017).	Режим доступа: http://www.intuit.ru/	Интернет-университет информационных технологий [Электронный ресурс].
Библиотека MSDN. [Электронный ресурс] : [официальный сайт] / Microsoft, -2014. - Режим доступа: http://msdn.microsoft.com/library/ свободный. - Загл. с экрана (дата обращения: 17.02.2017).	Режим доступа: http://msdn.microsoft.com/library/	Библиотека MSDN. [Электронный ресурс].

Таблица 3 – Дополнительная литература

Библиографическое описание* издания (автор, наименование, вид, место и год издания, кол.стр.)	Кол.экз. в библ. НовГУ	Наличие в ЭБС
1 Айвазян С. А. Прикладная статистика и основы эконометрики = Appliedstatisticsandessentialsfoeconometrics : учеб.для вузов / С. А. Айвазян, В. С. Мхитарян ; Гос. ун-т Высш. шк. экономики. - М. : ЮНИТИ, 1998. - 1022 с. : ил. - Библиогр.: с. 960-961.	3	
2 Грешилов А. А. Некорректные задачи цифровой обработки информации и сигналов / А. А. Грешилов. - 2-е изд., доп. - М. : Логос, 2009. - 358, [2] с. : ил. - Библиогр.: с. 345-358.	1	
3 Теребиж В.Ю. Введение в статистическую теорию обратных задач. - М. :Физматлит, 2005. - 375,[1]с.	1	

СОГЛАСОВАНО

НБ НовГУ:

должность

подпись

расшифровка