Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Кубанский государственный университет»
Факультет компьютерных технологий и прикладной математики

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.ДВ.03.03 МЕТОДЫ АНАЛИЗА ДАННЫХ

Направление подготовки **01.04.02 Прикладная математика и информатика**

Направленность (профиль) Математическое моделирование

Программа подготовки академическая

Форма обучения очная

Квалификация (степень) выпускника магистр

Краснодар 2017

Рабочая программа дисциплины «МЕТОДЫ АНАЛИЗА ДАННЫХ» составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки **01.04.02 Прикладная математика и информатика** (уровень магистратуры), утвержденным приказом Министерства образования и науки Российской Федерации № 911 от 28 августа 2015 г.

Программу составил:

Павлова А.В., д-р физ.-мат. наук, доцент, проф. кафедры математического моделирования КубГУ _____

Рабочая программа дисциплины «Методы анализа данных» утверждена на заседании кафедры математического моделирования
протокол № 16 « 21 » июня 2017 г.

Заведующий кафедрой математического моделирования акад. РАН, д-р физ.-мат. наук, проф. Бабешко В.А. _____

Утверждена на заседании учебно-методической комиссии факультета компьютерных технологий и прикладной математики
протокол № 4 « 29» июня 2017 г.

Председатель УМК факультета
канд. физ.-мат. наук, доцент Малыхин К.В _____

Рецензенты:

Калайдин Е.Н., д-р физ.-мат. наук, зав. кафедрой «Математика и информатика» Финансового университета при Правительстве РФ (Краснодарский филиал)

Ратнер С.В., д-р экон. наук, ведущий научный сотрудник Института проблем управления РАН им. В.А. Трапезникова

1 Цели и задачи изучения дисциплины

1.1 Цель освоения дисциплины

Цели изучения дисциплины определены государственным образовательным стандартом высшего образования и соотнесены с общими целями ООП ВО по направлению подготовки «Прикладная математика и информатика» (уровень магистратуры), в рамках которого преподается дисциплина.

Данная дисциплина ставит своей **целью** изучение алгоритмов анализа и интерпретации данных, формирование практических навыков использования современных программных средств решения задач анализа и интерпретации данных, формулировки гипотез об их структуре.

Процесс освоения данной дисциплины направлен на получения необходимого объема теоретических знаний, отвечающих требованиям ФГОС ВО и обеспечивающих успешное проведение магистром научно-исследовательской и проектно-производственной профессиональной деятельности. Цели дисциплины соответствуют следующим формируемым компетенциям: ПК-4, ПК-7, ПК-11.

1.2 Задачи дисциплины

Основные **задачи** дисциплины:

– изучение методов прикладного анализа данных;

– освоение компьютерных средств статистического анализа;

– приобретение практических навыков применения статистических методов для анализа данных и прогнозирования стохастических процессов.

1.3 Место дисциплины в структуре образовательной программы

Дисциплина «Методы анализа данных» относится к дисциплинам по выбору вариативной части Блока 1 "Дисциплины (модули)" подготовки магистра, базируется на знаниях, полученных по стандарту высшего образования.

Место курса в профессиональной подготовке магистра определяется ролью методов анализа и интерпретации данных в формировании высококвалифицированного специалиста в любой области знаний, использующей математические модели. Так как моделирование многих процессов и явлений, зависящих от большого количества параметров, неизбежно вызывает трудности, связанные с выявлением структуры взаимосвязей этих параметров, и производится на основе анализа стохастической информации, требующей от специалиста знаний и навыков применения методов анализа данных, дисциплина «Методы анализа данных» является важным звеном в обеспечении магистра знаниями, позволяющими прикладнику успешно вести профессиональную деятельность в сфере разработки математических моделей, а также обеспечивать полный цикл процесса моделирования.

Имеется логическая и содержательно-методическая взаимосвязь с другими частями ООП ВО. Дисциплина «Методы анализа данных» связана с дисциплинами базового цикла и другими дисциплинами вариативной части. Данный курс наиболее тесно связан с курсами: математические методы представления и анализа моделей, дискретные и вероятностные математические модели.

Необходимым требованием к «входным» знаниям, умениям и опыту деятельности обучающегося при освоении данной дисциплины, приобретенным в результате освоения предшествующих дисциплин является уверенное владение материалом курсов «Линейная алгебра», «Дискретная математика», «Теория вероятности и математическая статистика».

Программа определяет общий объем знаний, позволяющий сформировать у студента основы теоретических знаний по вопросам методики применения методов анализа данных и практических навыков применения статистических методов.

1.4 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы.

В результате изучения курса «методы анализа данных» студент должен овладеть:

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	**уметь**	**владеть**
1.	ПК-4	способностью разрабатывать и анализировать концептуальные и теоретические модели решаемых задач проектной и производственно-технологической деятельности	– задачи анализа данных; – подходы к анализу и интерпретации данных, получаемых с помощью информационно-измерительных систем. – основные принципы и методику анализа данных, принципы сбора и подготовки исходных данных; – основные принципы моделирования неопределённости	– ставить задачи интерпретации данных; – использовать основные методы анализа временных рядов; – применять основные методы и алгоритмы интеллектуального анализа.	– методами классификации данных; – методами корреляционного и дисперсионного анализа; – методами прогнозирования
2.	ПК-7	способностью разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов	– способы использования современных методов анализа для решения научных и практических задач	– самостоятельно выбрать метод анализа данных задачи – использовать современные теории для выбора метода.	– методами снижения размерности данных.
3.	ПК-11	способностью разрабатывать аналитические обзоры состояния области прикладной математики и информационных технологий	– современные тенденции развития методов анализа данных.	– применять изученные методы анализа данных при решении реальных практических задач; – содержательно интерпретировать и обосновывать	– навыками использования современных программных средств анализа данных.

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	**уметь**	**владеть**
				полученные результаты.	

Процесс освоения дисциплины «Методы анализа данных» направлен на получения необходимого объема знаний, отвечающих требованиям ФГОС ВО и обеспечивающих успешное ведение магистром научно-исследовательской, проектной и производственно-технологической деятельности.

Так как многие процессы и явления зависят от большого количества характеризующих их параметров, что обуславливает трудности, связанные с выявлением структуры взаимосвязей этих параметров, моделирование таких процессов производится на основе анализа стохастической, неполной информации, требующей от специалиста знаний и навыков применения методов анализа данных.

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 академических часов (из них 32 аудиторных). Курс «Методы анализа данных» состоит из лекционных и лабораторных занятий, сопровождаемых регулярной индивидуальной работой преподавателя со студентами в процессе самостоятельной работы. В конце семестра проводится экзамен. Программой дисциплины предусмотрены 16 часов лекционных и 16 часов лабораторных занятий, а также 49 часов самостоятельной работы и 27 часов подготовки к экзамену.

Знания и навыки, получаемые обучающимися в результате изучения дисциплины, необходимы для подготовки к работе с реальными данными в ходе ведения ими профессиональной деятельности.

Вид учебной работы	Всего часов	Семестр (часы)
		1
Контактная работа (всего)	**32,3**	**32,3**
В том числе:		
Занятия лекционного типа	16	16
Занятия семинарского типа (семинары, практические занятия)	–	–
Лабораторные занятия	16	16
Иная контактная работа:		
Контроль самостоятельной работы (КСР)	–	–
Промежуточная аттестация (ИКР)	0,3	0,3
Самостоятельная работа (всего)	**49**	**49**
В том числе:		

Вид учебной работы		Всего часов	Семестр (часы)
			1
Курсовая работа		–	–
Проработка учебного (теоретического) материала		25	25
Подготовка к текущему контролю		24	24
Контроль: экзамен			
Подготовка к экзамену		26,7	26,7
Общая трудоемкость	**час.**	**108**	**108**
	в том числе контактная работа	**32,3**	**323**
	зач. ед	**3**	**3**

2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины. Разделы дисциплины, изучаемые в 1 семестре

№	Наименование разделов	Количество часов				
		Всего	Аудиторная работа		Внеаудиторная работа	
			Л	ЛР	контроль	СРС
1	Анализ данных. Основные понятия и проблемы	8	2	2	2	2
2	Непараметрические методы классификации.	14	2	2	4	6
3	Классификация данных на основе статистических моделей	14	2	2	4	6
4	Кластерный анализ	14	2	2	4	6
5	Анализ матриц исходных данных	10	2	2	2	4
6	Методы снижения размерностей данных	10	2	2	2	4
7	Методы прогнозирования временных рядов	14	2	2	4	6
8	Системы Data Mining. в задачах анализа и интерпретации данных	23,7	2	2	4,7	15
Промежуточная аттестация (ИКР)		0,3	–	–	–	–
Итого:		**108**	**16**	**16**	**26,7**	**49**

Примечание: Л – лекции, ЛР – лабораторные занятия, СРС – самостоятельная работа студента.

2.3 Содержание разделов дисциплины:

№	Наименование раздела	Содержание раздела	Форма текущего контроля
1	2	3	4
1.	Анализ данных. Основные понятия	Проблема обработки данных. Матрица данных. Гипотезы компактности и скрытых факторов.	Опрос по результатам

№	Наименование раздела	Содержание раздела	Форма текущего контроля
1	2	3	4
	и проблемы	Структура матрицы данных и задачи обработки. Матрица объект-объект и признак-признак. Расстояние и близость. Измерение признаков. Отношения и их представление. Основные проблемы измерений. Основные типы шкал. Проблема адекватности. Основные задачи анализа и интерпретации данных	лабораторной работы
2.	Непараметрические методы классификации. Непараметрическое обучение дискриминантных функций.	Решающие поверхности и дискриминантные функции. Линейные дискриминантные функции класссификатор по минимуму расстояния. Линейная разделимость. Кусочно-линейные дискриминантные функции. Нелинейные дискриминантные функции. Потенциальные функции как дискриминантные функции. Пространство весов. Процедуры обучения с коррекцией ошибок: правило с фиксированным приращением, правило абсолютной коррекции, частично корректирующее правило. Обобщенные градиентные методы. Персептронный критерий. Процедуры обучения на основе минимальной среднеквадратичной ошибки: псевдоинверсный метод, метод Хо-Кашьяпа.	Опрос по результатам индивидуального задания
3.	Классификация данных на основе статистических моделей	Функция потерь. Байесовская дискриминантная функция. Принятие решение по максимуму правдоподобия. Оптимальная дискриминантная функция для нормально распределенных образов. Дискриминантная функция Фишера. Множественный дискриминантный анализ. Пошаговый дискриминантный анализ. Ошибки классификации. Примеры построения статистических дискриминантных функций для различных статистических нескольких моделей данных. Обучение для статистических дискриминантных функций. Оценки максимального правдоподобия, байесовские оценки. Непараметрическое оценивание. Метод непараметрического оценивания на основе К-ближайшего соседства.	Опрос по результатам лабораторных работ
4.	Кластерный анализ	Основные типы задач кластер-анализа. Меры подобия и функции расстояния. Выбор критерия кластеризации. Кластерные методы, основанные на евклидовой метрике. Иерархическая кластеризация. Метод К-внутригрупповых средних. Использование методов теории графов в задачах кластеризации. Кластеризация на основе анализа плотностей вероятностей	Опрос по результатам самостоятельной работы

№	Наименование раздела	Содержание раздела	Форма текущего контроля
1	2	3	4
5.	Анализ матриц исходных данных	Метод главных компонент. Корреляционная матрица и ее основные свойства. Собственные векторы и собственные числа корреляционной матрицы. Приведение корреляционной матрицы к диагональной форме. Геометрическая интерпретация главных компонент на плоскости.	Опрос по результатам лабораторных работ
6.	Методы снижения размерностей данных	Модели факторного анализа. Оценка факторных нагрузок методом максимального правдоподобия и центроидным методом. Вращение факторов и их интерпретация. Использование кластеризации признаков для снижения размерности. Многомерное шкалирование (МИ). Метрический и неметрический подход к МИ. Методы ортогонального проектирования. Нелинейные методы МИ. Многомерное шкалирование неметрических данных. Многомерные развертки.	Подготовка обзоров программных средств, опрос по результатам лабораторных работ
7.	Методы прогно-зирования вре-менных рядов	Классификация методов прогнозирования. Оценивание трендов. Методы скользящего среднего. Экспоненциальное сглаживание. Регрессионный анализ и прогнозирование. Линейные параметрические модели временных рядов. Методы оценивания моделей авторегрессии, скользящего среднего и смешанных моделей. Сезонные модели. Прогно-зирование на основе параметрических моделей. Прогнозирование с использованием нейронных сетей.	Опрос по результатам лабораторных работ
8.	Системы Data Mining. в задачах анализа и интерпретации данных	Понятие об интеллектуальных системах анализа и интерпретации данных. Data Mining -системы извлечения новых знаний из данных. Типы систем Data Mining -предметно-ориентированные аналитические системы, статистические пакеты, нейронные сети, деревья решений, обнаружение логических закономер-ностей, генетические алгоритмы, системы визуализации многомерных данных.	Подготовка презентации. Защита индивидуального задания

2.3.1 Занятия лекционного типа

Раздел 1. Проблема обработки данных. Матрица данных. Гипотезы компактности и скрытых факторов. Структура матрицы данных и задачи обработки. Матрица объект-объект и признак-признак. Расстояние и близость. Измерение признаков. Отношения и их представление. Основные проблемы измерений. Основные типы шкал. Проблема адекватности. Основные задачи анализа и интерпретации данных (2 ч.).

Раздел 2. Решающие поверхности и дискриминантные функции. Линейные дискриминантные функции класссификатор по минимуму расстояния. Линейная разделимость. Кусочно-линейные дискриминантные функции. Потенциальные функции как дискриминантные функции. Пространство весов. Процедуры обучения с коррекцией ошибок: правило с фиксированным приращением, правило абсолютной коррекции, частично корректирующее правило. Обобщенные градиентные методы. Процедуры обучения на основе минимальной среднеквадратичной ошибки: псевдоинверсный метод, метод Хо-Кашьяпа. (2.ч.).

Раздел 3. Функция потерь. Байесовская дискриминантная функция. Принятие решение по максимуму правдоподобия. Оптимальная дискриминантная функция для нормально распределенных образов. Дискриминантная функция Фишера. Множественный дискриминантный анализ. Пошаговый дискриминантный анализ. Ошибки классификации. Примеры построения статистических дискриминантных функций для различных статистических нескольких моделей данных. Обучение для статистических дискриминантных функций. Оценки максимального правдоподобия, байесовские оценки. Непараметрическое оценивание. Метод непараметрического оценивания на основе К-ближайшего соседства. (2 ч.).

Раздел 4. Основные типы задач кластерного анализа. Меры подобия и функции расстояния. Выбор критерия кластеризации. Кластерные методы, основанные на евклидовой метрике. Иерархическая кластеризация. Метод К-внутригрупповых средних. Использование методов теории графов в задачах кластеризации. Кластеризация на основе анализа плотностей вероятностей (2.ч.).

Раздел 5. Метод главных компонент. Корреляционная матрица и ее основные свойства. Собственные векторы и собственные числа корреляционной матрицы. Приведение корреляционной матрицы к диагональной форме. Геометрическая интерпретация главных компонент на плоскости. (2 ч.).

Раздел 6. Модели факторного анализа. Оценка факторных нагрузок методом максимального правдоподобия и центроидным методом. Вращение факторов и их интерпретация. Использование кластеризации признаков для снижения размерности. Многомерное шкалирование (МИ). Метрический и неметрический подход к МИ. Методы ортогонального проектирования. Нелинейные методы МИ. Многомерное шкалирование неметрических данных. Многомерные развертки. (2 ч.).

Раздел 7. Классификация методов прогнозирования. Оценивание трендов. Методы скользящего среднего. Экспоненциальное сглаживание. Регресссионный анализ и прогнозирование. Линейные параметрические модели временных рядов. Методы оценивания моделей авторегрессии, скользящего среднего и смешанных моделей. Сезонные модели. Прогнозирование на основе параметрических моделей. Прогнозирование с использованием нейронных сетей (2 ч.).

Раздел 8. Понятие об интеллектуальных системах анализа и интерпретации данных. Data Mining-системы извлечения новых знаний из данных. Типы систем Data Mining-предметно-ориентированные аналитические системы, статистические пакеты, нейронные сети, деревья решений, обнаружение логических закономерностей, генетические алгоритмы, системы визуализации многомерных данных. (2 ч.).

2.3.2 Занятия семинарского типа

Учебный план не предусматривает занятий семинарского типа по дисциплине «Методы анализа данных».

2.3.3 Лабораторные занятия

№ работы	№ раздела дисциплины	Наименование лабораторных работ	Форма текущего контроля
1	1	Методы и пакеты статистической обработки данных. Знакомство с программным обеспечением. Организация таблицы данных в пакетах Statistica, Statgraphics, SPSS.	Отчет по ЛР
2	3	Предварительный анализ данных с использованием пакета программ Statistica	Отчет по ЛР
3	4	Изучение методов дискриминантного анализа с использованием пакета программ Statistica	Отчет по ЛР
4	5	Изучение методов кластер-анализа с использованием пакета программ Statistica	Отчет по ЛР
5	6	Изучение методов факторного-анализа с использованием пакета программ Statistica	Отчет по ЛР
6	7	Классификация данных и изучение методов снижения размерности данных	Отчет по ЛР
7	8	Изучение методов прогнозирования временных рядов с использованием пакета программ Statistica	Отчет по ЛР
8	9	Использование технологий Data Mining	Отчет по ЛР

2.3.4 Примерная тематика курсовых работ (проектов)

Учебный план не предусматривает курсовых работ по дисциплине «Методы анализа данных».

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Подготовка к текущему контролю, подготовка индивидуальных зада–ний	1. Боровиков В.П. Популярное введение в современный анализ данных в системе STATISTICA. М.: Горячая линия-Телеком, 2013. 288 с. (+электронный ресурс https://e.lanbook.com/book/11828#authors). 2. Халафян А.А. Промышленная статистика: контроль качества, анализ процессов, планирование экспериментов в пакете STATISTICA. М.: URSS: Книжный дом "ЛИБРОКОМ", 2013. 380 с.

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в печатной форме увеличенным шрифтом,
– в форме электронного документа.
Для лиц с нарушениями слуха:
– в печатной форме,
– в форме электронного документа.
Для лиц с нарушениями опорно-двигательного аппарата:
– в печатной форме,
– в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Целью самостоятельной работы является углубление знаний, полученных в результате аудиторных занятий, выработка навыков индивидуальной работы, закрепление навыков, сформированных во время лабораторных занятий.

Самостоятельная работа подразумевает подготовку студентов к лабораторным занятиям, а также подготовку индивидуального итогового задания по отдельной теме согласно учебной программе дисциплины.

Самостоятельной работе по подготовке индивидуального задания отводится особое место. Преподавателем проводятся консультации, которые студент может посещать по желанию.

Наименование работы	**Кол-во часов**	**Форма контроля**
Проработка теоретического материала	**25**	Тематические сообщения на лабораторных занятиях
Подготовка к лабораторным занятиям (в том числе подготовка и защита итогового задания)	**24**	Обсуждение тематических вопросов на лабораторных занятиях, защита итогового задания
Всего часов самостоятельной работы	**49**	

3. Образовательные технологии

В соответствии с требованиями ФГОС ВО по направлению подготовки магистров программа по дисциплине «Методы анализа данных» предусматривает использование в учебном процессе следующих образовательные технологий и методов формирования компетенций: выполнение конкретных технических упражнений, знакомство с конкретными программными продуктами анализа и интерпретации данных, постановка и решение проблемных задач и т.д. Для реализации технологии коллективного взаимодействия отдельное занятие организуется в форме презентаций итоговых заданий с элементами дискуссий.

На лекциях используется интерактивная подача материала с мультимедийной системой. Лабораторные занятия предусматривают компьютерные эксперименты. Такое сочетание позволяет оптимально использовать отведенное время и раскрывать логику и содержание дисциплины.

<table>
<tr><th>Семестр</th><th>Вид занятия</th><th colspan="2">Используемые интерактивные образовательные технологии</th><th>Общее количество часов</th></tr>
<tr><td rowspan="5">1</td><td rowspan="4">Л</td><td colspan="2">Слайд-лекции. Представление индивидуальных заданий, обсуждение.</td><td>4</td></tr>
<tr><td>№</td><td>Тема</td><td>количество часов</td></tr>
<tr><td>1</td><td>Введение в анализ данных</td><td>2</td></tr>
<tr><td>2</td><td>Системы Data Mining. в задачах анализа и интерпретации данных</td><td>2</td></tr>
<tr><td>ЛР</td><td colspan="2">Компьютерные занятия в режимах взаимодействия «преподаватель – студент» и «студент – студент»</td><td>4</td></tr>
<tr><td colspan="4">Итого:</td><td>8</td></tr>
</table>

Цель ***лекции*** – обзор методов анализа и интерпретации данных, знакомство с проблемами и подходами.

Цель ***лабораторного занятия*** – научить применять теоретические знания при решении и исследовании конкретных задач. Лабораторные занятия проводятся в компьютерных классах, при этом практикуется работа в группах.

В ходе лабораторных занятий студент должен приобрести навыки решения типичных задач; применения программных систем, предназначенных для анализа данных, а также тестирования программных модулей на модельных данных; использования современных визуальных методов анализа данных и применения их для статистического вывода и формулировки гипотез о структуре данных; строить адекватные адаптивные модели для прогнозирования реальных данных и проводить по ним прогнозирование; осуществлять проверку адекватности и точности построенных моделей; самостоятельно принимать решения в задачах по выбору методов анализа в практических ситуациях.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Учебная деятельность проходит в соответствии с графиком учебного процесса. Процесс самостоятельной работы контролируется во время аудиторных занятий и индивидуальных консультаций.

Целью текущего контроля знаний является проверка ритмичности работы обучаемых, оценка усвоения теоретического и практического материала, а также приобретенных знаний, умений и навыков.

Текущий контроль осуществляется на каждом лабораторном занятии преподавателем, проводящим эти занятия. Текущий контроль выполнения самостоятельной работы осуществляется преподавателем на консультациях.

Фонд оценочных средств дисциплины состоит из средств текущего контроля (см. список лабораторных работ, тем и вопросов для самоподготовки, тем рефератов) и промежуточной аттестации (экзамена).

Обязательным условием допуска студента к экзамену является успешное выполнение индивидуального задания и лабораторных работ.

Оценка успеваемости осуществляется по результатам: самостоятельного

выполнения лабораторных работ, устного опроса при сдаче выполненных лабораторных заданий и защиты индивидуального задания, ответа на экзамене.

Экзамен проводится устно, ему предшествует защита индивидуальных заданий, выполнение которых требует знания как теоретического материала, так и практических навыков решения. Индивидуальное задание является важным элементом технологии адаптивного обучения.

Соответствие компетенций, формируемых при изучении дисциплины, и видов занятий

Перечень компетенций	Виды занятий			Формы контроля
	Л.	Лаб.	СРС	
ПК-4	+	+	+	– Подготовка реферата; – Защита индивидуального задания
ПК-7, ПК-11		+	+	– Подготовка реферата; – Опрос по результатам выполнения лабораторных заданий; – Опрос по результатам самостоятельной работы; – Подготовка презентации и защита индивидуального задания

4.1 Фонд оценочных средств для проведения текущего контроля

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующие этапы формирования компетенций в процессе освоения образовательной программы

Примерные задания на лабораторные работы

1. Методы и пакеты статистической обработки данных. Знакомство с программным обеспечением. Организация таблицы данных в пакетах Statistica, Statgraphics, SPSS.Растяжение плоскости с эллиптическим отверстием
2. Предварительный анализ данных с использованием пакета программ Statistica Напряжение в вершине трещины нормального отрыва в условиях плоского напряженного и плоского деформированного состояний в идеально пластическом материале
3. Изучение методов дискриминантного анализа с использованием пакета программ Statistica
4. Изучение методов кластер-анализа с использованием пакета программ Statistica
5. Изучение методов факторного-анализа с использованием пакета программ Statistica
6. Классификация данных и изучение методов снижения размерности данных
7. Изучение методов прогнозирования временных рядов с использованием пакета программ Statistica
8. Использование технологий Data Mining

Примерные задания для устного опроса

1. В чем состоят цели многомерного анализа данных?
2. Дайте формальное описание данных, используемое в многомерном статистическом анализе, и опишите представление данных в электронном виде.
3. Определите основные шкалы измерений в многомерных выборках.
4. Как произвести преобразование данных из одной шкалы в другую?
5. Определите понятие диапазон значений признака и способы определения диапазона значений.
6. Назовите свойства эмпирических теорий.
7. Назовите критерии информативности признаков.
8. Охарактеризуйте меры близости между признаками, измеренными в шкале наименований.
9. Охарактеризуйте меры близости между признаками, измеренными в шкале порядка.
10. Назовите основные характеристики многомерной выборки.
11. Как рассчитываются основные характеристики многомерной выборки в предварительном анализе данных?
12. Определите средства визуализации данных в предварительном анализе данных.
13. Для чего используются модельные данные при решении задач многомерного анализа?
14. В чем состоит идея метода неравномерной рулетки?
15. В чем состоит идея метода отбраковки?
16. Определите способы моделирования случайных чисел выборок в Excel.
17. Определите принципы моделирования многомерного нормального распределения?
18. Определите проблему робастности в анализе данных.
19. Что понимается под грубыми ошибками и каковы последствия их проявлений в статистической совокупности?
20. Назовите известные вам критерии робастности.
21. Дайте характеристику алгоритмов определения «подозрительных» объектов выборки.
22. Каковы последствия отсутствия значений в многомерной выборке и как с ними бороться.
23. Какие существуют способы обработки грубых ошибок?
24. Назовите виды нормировки данных.
25. Что понимается под статистической гипотезой и какие характерные признаки для нее существуют?
26. Какого рода ошибки могут допускаться и чем определяется достоверность выводов при проверке статистических гипотез?
27. Определите предпосылки применения критерия Фишера.
28. Определите предпосылки применения критерия Стьюдента.
29. Как подсчитывается сумма рангов в критерии Вилксона?
30. Какое распределение используется в критерии Вилксона?

Примерный вариант внеаудиторного задания

1. Используя файл с данными наблюдений об уровне рек в крае monitoring.xls, проведите сравнительный корреляционно-регрессионный анализ данных в целом по региону и по указанному району, а также по этому району с учетом его площади и количества датчиков.

2. Проведите классификацию районов по уровню угрозы паводка.

3. Проведите факторный анализ исходных данных об уровне воды в реках.

Примерные темы рефератов

1. Методы анализа знаний.

2. Теория измерений.
3. Анализ данных, знаний и структур в системах искусственного интеллекта.
4. Прогнозирование многомерных временных рядов.
5. Методы анализа структурных объектов.
6. Программные продукты для интеллектуального анализа.
7. Методы классификации признаков. Кластерный анализ и деревья классификации.
8. Методы планирования эксперимента.
9. Обучающиеся генетические алгоритмы прогнозирования.
10. Экспертные системы (назначение, модели представления знаний, достоинства).

4.2 Фонд оценочных средств для проведения промежуточной аттестации

Основные требования к результатам освоения дисциплины представлены в таблице в виде признаков сформированности компетенций. Требования формулируются по двум уровням: пороговый и повышенный и в соответствии со структурой, принятой в ФГОС ВО: знать, уметь, владеть.

Название компетенции (или ее части)	**Структура компетенции**	**Основные признаки сформированности компетенции**
ПК-4 способностью разрабатывать и анализировать концептуальные и теоретические модели решаемых задач проектной и производственно-технологической деятельности	Знать: основные понятия и задачи анализа данных, подходы к их решению	Знает основные понятия и задачи анализа данных, подходы к их решению
		Знает основные принципы и методику анализа данных, принципы сбора и подготовки исходных данных
		Знает теоретическую основу методов анализа данных
	Уметь: ставить задачи интерпретации данных.	Умеет ставить задачи интерпретации данных
		Умеет применять основные методы и алгоритмы интеллектуального анализа
	Владеть: основными методами анализа данных в прикладных программах	Владеет методами классификации данных, корреляционного и дисперсионного анализа разнотипных данных,
		Владеет методом прогнозирования в прикладных программах
ПК-7 способностью разрабатывать и оптимизировать бизнес-планы научно-прикладных проектов	Знать способы использования современных методов анализа для решения научных и практических задач	Знает способы использования современных методов для решения научных и практических задач,
		Знает принципы выбора методов и средств изучения математической модели
	Уметь самостоятельно выбрать	Умеет использовать современные теории для выбора метода.

Название компетенции (или ее части)	Структура компетенции	Основные признаки сформированности компетенции
	метод анализа данных задачи	Умеет провести сравнение методов, самостоятельно выбирать метод анализа данных задачи, доказывать его эффективность
	Владеть методами снижения размерности данных.	Владеет навыками оптимизации обработки большого объёма данных
ПК-11 способностью разрабатывать аналитические обзоры состояния области прикладной математики и информационных технологи	Знать основные алгоритмы анализа данных, группировать первичные экспериментальные данные, представлять их в графической форме	Знает основные алгоритмы анализа данных, основные математические и статистические методы исследования, используемые при анализе данных
	Уметь грамотно обработать и извлечь из собранных данных информацию, группировать первичные экспериментальные данные, представлять их в графической форме	Умеет грамотно обработать и извлечь из собранных данных максимум информации
		Умеет систематизировать и анализировать данные
		Умеет проводить факторный и кластерный анализ данных
		Умеет прогнозировать будущее поведение некоторого временного ряда
		Умеет группировать первичные экспериментальные данные, представлять их в графической форме
	Владеть инструментальными средствами для решения конкретной задачи на компьютере	Владеет инструментальными средствами для решения конкретной задачи на компьютере, содержательной интерпретации результатов

Перечень вопросов, выносимых на экзамен

1. Основные задачи анализа и интерпретации данных.
2. Матрица данных. Гипотезы компактности и скрытых факторов.
3. Решающие поверхности и дискриминантные функции.
4. Функция потерь. Байесовская дискриминантная функция.
5. Приведение корреляционной матрицы к диагональной форме.
6. Методы статистического оценивания и сравнения выборок.
7. Непараметрические методы проверки однородности выборок.
8. U-критерий Манна-Уитни (Вилксона).
9. Дисперсионный анализ.
10. Кластерные методы, основанные на евклидовой метрике.
11. Использование кластеризации признаков для снижения размерности.
12. Многомерное шкалирование.
13. Регрессионный анализ и прогнозирование.
14. Линейные параметрические модели временных рядов.

15. Прогнозирование с использованием нейронных сетей.
16. Проверка гипотезы с помощью F-критерия
17. Методы обработки ранговых данных.
18. Коэффициент ассоциации Юла. Ранговая корреляция. Критерий знаков. Методы проведения экспертных исследований и анализ данных оценок экспертов.
19. Методы моделирования случайных величин. Метод неравномерной рулетки. Метод отбраковки.
20. Быстрый способ моделирования одномерного нормального распределения.
21. Моделирование многомерного нормального распределения.
22. Определение Data Mining. Задачи анализа данных.
23. Методы и модели Data Mining.
24. Стандарты Data Mining: Стандарт CWM; стандарт CRISP; стандарт PMML.
25. Генетические алгоритмы.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в печатной форме увеличенным шрифтом,

– в форме электронного документа.

Для лиц с нарушениями слуха:

– в печатной форме,

– в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме,

– в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Методические рекомендации к сдаче экзамена

Экзамен является заключительным этапом процесса формирования компетенции студента при изучении дисциплины или ее части и имеет целью проверку и оценку знаний студентов по теории и применению полученных знаний, умений и навыков при решении практических задач. Экзамены проводятся по расписанию, сформированному учебным отделом и утвержденному проректором по учебной работе, в сроки, предусмотренные календарным графиком учебного процесса. Расписание экзаменов доводится до сведения студентов не менее чем за две недели до начала экзаменационной сессии. Экзамены принимаются преподавателями, ведущими лекционные занятия. В отдельных случаях при большом количестве групп у одного лектора или при большой численности группы с разрешения заведующего кафедрой допускается привлечение в помощь основному лектору преподавателя, проводившего практические занятия в группах.

Экзамены проводятся в устной форме. Экзамен проводится только при предъявлении студентом зачетной книжки и при условии выполнения всех контрольных мероприятий, предусмотренных учебным планом и рабочей программой по изучаемой дисциплине. Студентам на экзамене предоставляется право выбрать один из билетов. Время подготовки к ответу составляет 60 минут. По истечении установленного времени студент должен ответить на вопросы экзаменационного билета и предоставить решение задач. Результаты экзамена оцениваются по четырехбалльной системе («отлично», «хорошо», «удовлетворительно», «неудовлетворительно») и заносятся в экзаменационную ведомость и зачетную книжку. В зачетную книжку заносятся только положительные оценки.

Критерии выставления оценок

Оценка «*отлично*»:

– систематизированные, глубокие и полные знания по всем разделам дисциплины, а также по основным вопросам, выходящим за пределы учебной программы;

– точное использование научной терминологии систематически грамотное и логически правильное изложение ответа на вопросы;

– безупречное владение инструментарием учебной дисциплины, умение его эффективно использовать в постановке и решении научных и практических задач;

– выраженная способность самостоятельно и творчески решать сложные проблемы и нестандартные ситуации;

– полное и глубокое усвоение основной и дополнительной литературы, рекомендованной учебной программой по дисциплине;

– умение ориентироваться в теориях, концепциях и направлениях дисциплины и давать им критическую оценку, используя научные достижения других дисциплин;

–творческая самостоятельная работа на практических занятиях, активное участие в групповых обсуждениях, высокий уровень культуры исполнения заданий;

– высокий уровень сформированности заявленных в рабочей программе компетенций.

Оценка «*хорошо*»:

– достаточно полные и систематизированные знания по дисциплине;

– умение ориентироваться в основных теориях, концепциях и направлениях дисциплины и давать им критическую оценку;

– использование научной терминологии, лингвистически и логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;

– владение инструментарием по дисциплине, умение его использовать в постановке и решении научных и профессиональных задач;

– усвоение основной литературы, рекомендованной учебной программой по дисциплине;

– самостоятельная работа на практических занятиях, участие в групповых обсуждениях, средний уровень культуры исполнения заданий;

– средний уровень сформированности заявленных в рабочей программе компетенций.

Оценка «*удовлетворительно*»:

– достаточный минимальный объем знаний по дисциплине;

– усвоение основной литературы, рекомендованной учебной программой;

– умение ориентироваться в основных теориях, концепциях и направлениях по дисциплине и давать им оценку;

– использование научной терминологии, стилистическое и логическое изложение ответа на вопросы, умение делать выводы без существенных ошибок;

– владение инструментарием учебной дисциплины, умение его использовать в решении типовых задач;

– умение под руководством преподавателя решать стандартные задачи;

– работа под руководством преподавателя на практических занятиях, допустимый уровень культуры исполнения заданий;

– достаточный минимальный уровень сформированности заявленных в рабочей программе компетенций.

Оценка «*неудовлетворительно*»:

– фрагментарные знания по дисциплине;

– отказ от ответа;

– знание отдельных источников, рекомендованных учебной программой по дисциплине;

– неумение использовать научную терминологию;

– наличие грубых ошибок;

– низкий уровень культуры исполнения заданий;

–низкий уровень сформированности заявленных в рабочей программе компетенций.

5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

5.1 Основная литература:

1. Боровиков, В.П. Популярное введение в современный анализ данных в системе STATISTICA. М.: Горячая линия-Телеком, 2013. 288 с. [Электронный ресурс]. ‑ Режим доступа: https://e.lanbook.com/book/11828.

2. Крянев, А.В. Метрический анализ и обработка данных / А.В. Крянев, Г.В. Лукин, Д.К. Удумян. М.: Физматлит, 2012. 308 с. [Электронный ресурс]. - Режим доступа: https://e.lanbook.com/book/59523.

3. Халафян, А.А. Промышленная статистика: контроль качества, анализ процессов, планирование экспериментов в пакете STATISTICA. М.: Книжный дом "ЛИБРОКОМ", 2013. 380 с.

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечных системах.

5.2 Дополнительная литература:

1. Подкорытова О.А., Соколов М.В. Анализ временных рядов. СПб.: Юрайт, 2017. 266 с. [Электронный ресурс]. - Режим доступа: https://www.biblio-online.ru/book/7132122F-D176-4118-AD03-D43A9FA2FF86).

2. Халафян А.А. Математическая статистика с элементами теории вероятностей. STATISTICA 6 БИНОМ-Пресс, 2011. 491 с.

3. Халафян А. А. Статистический анализ данных. STATISTICA 6. М.: БИНОМ-Пресс, 2010. 522 с.

4. Чубукова И.А. Data Mining. М.: Интернет-Университет Информационных Технологий: БИНОМ. Лаборатория знаний, 2008. 382 с. + [Электронный ресурс]. – Режим доступа: http://biblioclub.ru/index.php?page=book&id=233055.

5.3. Периодические издания:

Не используются.

6. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Статистика http://statsoft.ru
2. Журнал «Машинное обучение и анализ данных» http://jmlda.org/ru/

7. Методические указания для обучающихся по освоению дисциплины

Анализ данных следует рассматривать как изучение методик практического применения теоретических методов, прежде всего методов математической статистики. Прикладной характер, изучаемой дисциплины подчеркивается ориентацией ее на конкретные программные продукты.

В процессе изучения дисциплины обучаемые готовят обзоры программных средств, предназначенных для проведения анализа данных. Обзор в обязательном порядке включает общие сведения о системе, информацию о ее назначении и компонентах, реализованных методы анализа и возможностях представления данных и результатов.

В рамках самостоятельной работы студент готовит индивидуальную работу. Каждый студент выполняет работу по одной теме. В конце семестра студент представляет работу в виде презентации и сдает отчет преподавателю.

Перечень вопросов для самоподготовки

1. Методы и пакеты статистической обработки данных. Знакомство с программным обеспечением. Организация таблицы данных в пакетах Statistica, Statgraphics, SPSS.
2. Виды графиков. Графический анализ данных.
3. Описательные статистики, точечные и интервальные оценки параметров генеральной совокупности.
4. Проверка статистических гипотез. Параметрические и непараметрические методы проверки.
5. Нелинейные дискриминантные функции.
6. Методы исследования взаимосвязей: корреляционный анализ.
7. Методы исследования взаимосвязей: анализ таблиц.
8. Дисперсионный анализ.
9. Регрессионный анализ.
10. Многомерный дисперсионный анализ.
11. Канонический корреляционный анализ.
12. Методы снижения размерности: факторный анализ.
13. Многомерное шкалирование.
14. Методы классификации многомерных наблюдений: кластерный анализ.
15. Методы классификации многомерных наблюдений: дискриминантный анализ.
16. Анализ многомерных наблюдений методом нейронных сетей.
17. Персептронный критерий.
18. Графическое представление результатов исследования.
19. Моделирование многомерного нормального распределения.
20. Способы генерации данных в EXCEL. Методы размножения выборок (бустреб-методы).
21. Методы вычисления устойчивых статистических оценок: Пуанкаре, Винзора, Хубера.

22. Генетические алгоритмы.
23. Определение Data Mining. Задачи анализа данных.
24. Методы и модели Data Mining.
25. Практическое применение Data Mining.
26. Средства Data Mining. Weka.
27. Стандарты Data Mining: Стандарт CWM; стандарт CRISP; стандарт PMML.

Поиск информации для ответов на вопросы для самостоятельной работы и выполнения заданий в некоторых случаях предполагает не только изучение основной учебной литературы, но и привлечение дополнительной литературы, а также использование ресурсов сети Интернет.

В рамках самостоятельной работы студент готовит реферативную работу, объемом не В рамках самостоятельной работы студент готовит реферат по выбранной теме. Каждый студент выполняет работу по одной теме.

Для написания реферата необходимо подобрать литературу. Общее количество литературных источников, включая тексты из Интернета, (публикации в журналах), должно составлять не менее 10 наименований. Учебники в литературные источники не входят.

Рефераты выполняют на листах формата А4. Страницы текста, рисунки, формулы нумеруют, рисунки снабжают подрисуночными надписями. Текст следует печатать шрифтом №14 с интервалом между строками в 1,5 интервала, без недопустимых сокращений. В конце работы приводят список использованных источников.

Обзор должен быть подписан магистрантом с указанием даты ее оформления.

Работы, выполненные без соблюдения перечисленных требований, возвращаются на доработку.

Выполненная магистрантом работа определяется на проверку преподавателю в установленные сроки. Если у преподавателя есть замечания, работа возвращается и после исправлений либо вновь отправляется на проверку, если исправления существенные, либо предъявляется на экзамене, где происходит ее защита.

Для приобщения обучаемых к поиску и исследовательской работе, для развития их творческого потенциала следует по возможности избегать прямого руководства работой обучающихся при выполнении ими тех или иных заданий, чаще выступать в роли консультанта, эксперта.

Предпочтительным является представление рефератов в форме конференции.

Примерные темы рефератов представлены в п. 4.1.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

8.1 Перечень информационных технологий

– Проверка индивидуальных заданий и консультирование посредством электронной почты.

– Использование электронных презентаций при проведении лекционных и лабораторных занятий.

– Использование статистических и математических пакетов при проведении лабораторных занятий.

8.2 Перечень необходимого лицензионного и свободного программного обеспечения

1. Операционная система MS Windows.
2. Интегрированное офисное приложение MS Office.
3. Программное обеспечение для организации управляемого коллективного и безопасного доступа в Интернет.
4. Математические пакеты Statistica, Matlab.

8.3 Перечень информационных справочных систем:

– Портал Федерального агентства по техническому регулированию и метрологии http://www.gost.ru;

– Электронная библиотечная система "Юрайт" (http://www.biblio-online.ru).

– Электронная библиотечная система "Университетская библиотека ONLINE" (http://www.biblioclub.ru).

– Электронная библиотечная система издательства "Лань" (http://e.lanbook.com).

– База данных Научной электронный библиотеки eLIBRARY.RU https://elibrary.ru/

– Глобальный интеллектуальный ресурс http://statistica.ru/local-portals/

– Базы данных и аналитические публикации «Университетская информационная система РОССИЯ» https://uisrussia.msu.ru/

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

№	Вид работ	Материально-техническое обеспечение дисциплины и оснащенность
1.	Лекционные занятия	Лекционная аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук), соответствующим программным обеспечением, а также необходимой мебелью (доска, столы, стулья). (аудитории: 129, 131, 133, А305, А307).
2.	Лабораторные занятия	Компьютерный класс, укомплектованный компьютерами с лицензионным программным обеспечением, необходимой мебелью (доска, столы, стулья). (аудитории: 101, 102, 106, 106а, 105/1, 107(2), 107(3), 107(5), А301).
3.	Групповые (индивидуальные) консультации	Аудитория для семинарских занятий, групповых и индивидуальных консультаций, укомплектованные необходимой мебелью (доска, столы, стулья). (аудитории: 129, 131).
4.	Текущий контроль, промежуточная аттестация	Аудитория для семинарских занятий, текущего контроля и промежуточной аттестации, укомплектованная необходимой мебелью (доска, столы, стулья) (аудитории: 129, 131, 133, А305, А307, 147, 148, 149, 150, 100С, А301б, А512), компьютерами с лицензионным программным обеспечением и выходом в интернет (106, 106а, А301)

№	Вид работ	Материально-техническое обеспечение дисциплины и оснащенность
5.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения, обеспеченный доступом в электронную информационно-образовательную среду университета, необходимой мебелью (доска, столы, стулья). (Аудитория 102а, читальный зал).

Осуществление учебного процесса предполагает наличие необходимого для реализации данной программы перечня материально-технического обеспечения: аудитории, оборудованные видеопроекционным оборудованием для презентаций (цифровой проектор, экран, ноутбук) и необходимой мебелью (доска, столы, стулья); компьютерные классы с компьютерной техникой с лицензионным программным обеспечением и необходимой мебелью (доска, столы, стулья) для проведения занятий.

Компьютерная поддержка учебного процесса по направлению 01.04.02 Прикладная математика и информатика обеспечивается по всем дисциплинам. Факультет компьютерных технологий и прикладной математики, оснащен компьютерными классами, установлена локальная сеть, все компьютеры факультета подключены к сети Интернет. Магистрантам доступны современные ПЭВМ, современное лицензионное программное обеспечение.

Магистранты и преподаватели вуза имеют постоянный доступ к электронному каталогу учебной, методической, научной литературе, периодическим изданиям и архиву статей.