

Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Оренбургский государственный университет»

Кафедра математических методов и моделей в экономике

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«Б.1.В.ОД.11 Многомерный статистический анализ»

Уровень высшего образования

БАКАЛАВРИАТ

Направление подготовки

01.03.02 Прикладная математика и информатика
(код и наименование направления подготовки)

Общий профиль

(наименование направленности (профиля) образовательной программы)

Тип образовательной программы

Программа академического бакалавриата

Квалификация

Бакалавр

Форма обучения

Очная

Год набора 2018

Рабочая программа рассмотрена и утверждена на заседании кафедры

Кафедра математических методов и моделей в экономике

наименование кафедры

протокол № 6 от "29" 01 2018 г.

Кафедра математических методов и моделей в экономике

наименование кафедры



А.Г. Реннер

расшифровка подписи

Исполнители:

Доцент кафедры ММиМЭ

должность



подпись

О.С. Чудинова

расшифровка подписи

должность

подпись

расшифровка подписи

СОГЛАСОВАНО:

Председатель методической комиссии по направлению подготовки

01.03.02 Прикладная математика и информатика

код наименование



личная подпись

И.П. Болодурина

расшифровка подписи

Заведующий отделом комплектования научной библиотеки



личная подпись

Н.Н. Грицай

расшифровка подписи

Уполномоченный по качеству факультета



личная подпись

Н.В. Лужнова

расшифровка подписи

№ регистрации _____

© Чудинова О.С., 2018

© ОГУ, 2018

1 Цели и задачи освоения дисциплины

Цель освоения дисциплины: овладение студентами методами многомерного анализа данных об экономических или естественнонаучных объектах, системах, процессах.

Задачи:

- изучение методов исследования взаимосвязей между показателями, характеризующими объекты в социально-экономических или естественнонаучных исследованиях;
- изучение методов распознавания образов и типологизации объектов;
- изучение методов оптимизации представления информации об объектах;
- освоение современных пакетов прикладных программ, реализующих алгоритмы многомерного анализа данных;
- приобретение навыков содержательной интерпретации результатов исследования.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к обязательным дисциплинам (модулям) вариативной части блока 1 «Дисциплины (модули)»

Пререквизиты дисциплины: *Б.1.Б.16 Теория вероятностей и математическая статистика, Б.1.Б.21 Методы оптимизации*

Постреквизиты дисциплины: *Б.2.В.П.3 Преддипломная практика*

3 Требования к результатам обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих результатов обучения

Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций	Формируемые компетенции
<u>Знать:</u> основные методы и алгоритмы многомерного статистического анализа, а также основы работы с программным обеспечением, реализующим эти методы и алгоритмы; <u>Уметь:</u> выбирать и применять инструментарий многомерного статистического анализа данных для решения возникающих на практике задач; <u>Владеть:</u> навыками работы с программным обеспечением, реализующим методы и алгоритмы многомерного статистического анализа данных, а также навыками программной реализации методов и алгоритмов для решения возникающих на практике задач.	ОПК-3 способностью к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям
<u>Знать:</u> инструментарий многомерного статистического анализа; возможности использования и сферы применения всех изученных методов и алгоритмов анализа данных; <u>Уметь:</u>	ОПК-4 способностью решать стандартные задачи профессиональной деятельности на основе информационной и

Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций	Формируемые компетенции
решать стандартные задачи профессиональной деятельности с помощью изученных методов и алгоритмов многомерного статистического анализа данных, в том числе применяя специализированное программное обеспечение, реализующее эти методы и алгоритмы; проводить анализ и давать интерпретацию полученным результатам; предлагать решения на основе полученных результатов; Владеть: навыками решения стандартных задач профессиональной деятельности с помощью изученных методов и алгоритмов многомерного статистического анализа данных, в том числе с применением специализированного программного обеспечения, реализующего эти методы и алгоритмы; навыками анализа и интерпретации полученных результатов; навыками разработки на основе полученных результатов решений или их элементов.	библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности
Знать: основные понятия и инструментарий многомерного статистического анализа данных, необходимые для критического переосмысления накопленного опыта, изменения вида и характера профессиональной деятельности; Уметь: используя методы и алгоритмы многомерного статистического анализа, критически переосмысливать накопленный опыт, изменять при необходимости вид и характер своей профессиональной деятельности; Владеть: навыками работы с программным обеспечением, реализующим методы и алгоритмы многомерного статистического анализа данных, для критического переосмысления накопленного опыта, изменения вида и характера профессиональной деятельности.	ПК-3 способностью критически переосмысливать накопленный опыт, изменять при необходимости вид и характер своей профессиональной деятельности

4 Структура и содержание дисциплины

4.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единиц (108 академических часов).

Вид работы	Трудоемкость, академических часов	
	7 семестр	всего
Общая трудоёмкость	108	108
Контактная работа:	34,25	34,25
Лекции (Л)	18	18
Лабораторные работы (ЛР)	16	16
Промежуточная аттестация (зачет, экзамен)	0,25	0,25
Самостоятельная работа: - выполнение индивидуального творческого задания (ИТЗ); - самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий); - подготовка к лабораторным занятиям).	73,75	73,75
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	зачет	

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Введение	8	2	-	-	6
2	Оценивание и сравнение параметров многомерной генеральной совокупности	12	2	-	2	8
3	Многомерный корреляционный анализ количественных признаков	12	2	-	2	8
4	Классическая линейная модель множественной регрессии	18	2	-	4	12
5	Методы кластерного анализа	20	2	-	4	14
6	Дискриминантный анализ	18	4	-	2	12
7	Метод главных компонент	20	4	-	2	14
	Итого:	108	18		16	74
	Всего:	108	18		16	74

4.2 Содержание разделов дисциплины

№ 1 Введение

Объект, предмет, задачи многомерного статистического анализа данных. Многомерная генеральная совокупность и ее закон распределения. Многомерный нормальный закон распределения. Представление исходной информации в многомерном анализе.

№ 2 Оценивание и сравнение параметров многомерной генеральной совокупности

Точечное оценивание параметров распределения многомерной генеральной совокупности. Построение доверительной области для вектора (части компонент вектора) математических ожиданий нормально распределенной генеральной совокупности. Построение доверительной области для вектора параметров в форме прямоугольного параллелепипеда. Проверка гипотезы о равенстве вектора математических ожиданий нормально распределенной генеральной совокупности стандарту. Проверка гипотезы об однородности распределения двух генеральных совокупностей.

№ 3 Многомерный корреляционный анализ количественных признаков

Теория корреляции: определение и свойства коэффициента корреляции и корреляционной матрицы, наилучшая линейная аппроксимация одной случайной величины остальными ($k-1$) компонентами случайного вектора, функция регрессии, остаточная, факторная дисперсии, корреляционное отношение, коэффициент детерминации, определение и свойства частного коэффициента корреляции. Многомерный корреляционный анализ: постановка задачи; оценка матрицы парных коэффициентов корреляции, частных коэффициентов корреляции, множественного коэффициента корреляции, коэффициента детерминации, функции регрессии; проверка гипотез о значимости характеристик связи и построение доверительных интервалов.

№ 4 Классическая линейная модель множественной регрессии

Функция регрессии, линейная модель множественной регрессии, условия Гаусса-Маркова, классическая линейная модель множественной регрессии (КЛММР). Метод наименьших квадратов (МНК) оценки параметров КЛММР. Выборочный коэффициент детерминации. Статистические свойства оценок параметров КЛММР. Мультиколлинеарность и методы ее устранения.

№ 5 Методы кластерного анализа

Постановка задачи классификации без обучения (непараметрический случай). Требования к мерам сходства и расстояниям между объектами. Метрики расчета расстояния между объектами. Способы расчета расстояния между классами объектов. Классификация задач кластерного анализа и основные типы кластер-процедур. Агломеративные методы кластерного анализа: метод одиночной

связи, метод полных связей, метод средней связи, метод Уорда. Дивизимные методы кластерного анализа. Итерационные методы кластерного анализа: метод К-средних. Функционалы качества разбиения.

№ 6 Дискриминантный анализ

Постановка задачи классификации с обучением (параметрический случай). Понятие класса. Функции потерь и вероятности неверной классификации. Построение оптимальных процедур классификации. Параметрический дискриминантный анализ в случае нормального закона распределения классов и его геометрическая интерпретация.

№ 7 Метод главных компонент

Постановка задачи снижения размерности признакового пространства, понятие меры информативности новой системы признаков. Необходимость и возможность снижения размерности признакового пространства. Определение 1-ой, 2-ой, ..., k-ой главных компонент. Оптимизационная задача для построения первой главной компоненты и процедура её решения. Основные числовые характеристики главных компонент. Матрица нагрузок и её свойства. Алгоритм оценки главных компонент. Матрица индивидуальных значений главных компонент.

4.3 Лабораторные работы

№ ЛР	№ раздела	Наименование лабораторных работ	Кол-во часов
1	2	Оценивание и сравнение параметров многомерной нормально распределенной генеральной совокупности	2
2	3	Многомерный корреляционный анализ количественных признаков	2
3	4	Классическая линейная модель множественной регрессии	4
4	5	Методы кластерного анализа	4
5	6	Параметрический дискриминантный анализ	2
6	7	Метод главных компонент	2
		Итого:	16

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

1. Методы и модели эконометрики [Электронный ресурс] : учебное пособие для студентов, обучающихся по программам высшего образования по направлениям подготовки 01.03.04 Прикладная математика, 38.04.01 Экономика, 38.03.05 Бизнес-информатика / О.И. Бантикова, В.И. Васянина, Ю.А. Жемчужникова, А.Г. Реннер, Е.Н. Седова, О.И. Стебунова, Л.М. Туктамышева, О.С. Чудинова /под ред. А. Г. Реннера. - Ч. 1. Анализ данных. - Оренбург : ОГУ, 2015.

2. Методы и модели эконометрики [Электронный ресурс] : учебное пособие для студентов, обучающихся по программам высшего образования по направлениям подготовки 01.03.04 Прикладная математика, 38.04.01 Экономика, 38.03.05 Бизнес-информатика / О.И. Бантикова, В.И. Васянина, Ю.А. Жемчужникова, А.Г. Реннер, Е.Н. Седова, О.И. Стебунова, Л.М. Туктамышева, О.С. Чудинова /под ред. А. Г. Реннера. - Часть 2. Эконометрика пространственных данных. - Оренбург : ОГУ, 2015.

5.2 Дополнительная литература

1. Айвазян, С.А. Прикладная статистика. Основы эконометрики: учебник для вузов: в 2 т. / С.А. Айвазян, В.С. Мхитарян. – М.: ЮНИТИ-ДАНА, 2001. – Т.1: Теория вероятностей и прикладная статистика. – 656 с.

2. Большаков, А. А. Методы обработки многомерных данных и временных рядов: учеб. пособие для вузов / А. А. Большаков, Р. Н. Каримов. - М.: Горячая линия - Телеком, 2007. - 522 с.
3. Крутиков, В.Н. Анализ данных: учебное пособие [Электронный ресурс] / В.Н. Крутиков, В.В. Мешечкин. - Кемерово, 2014. - 170 с. - Режим доступа: http://biblioclub.ru/index.php?page=book_view&book_id=278426
4. Математическое моделирование: исследование социальных, экономических и экологических процессов (региональный аспект) [Комплект] : учебное пособие для студентов, обучающихся по программам высшего профессионального образования по экономическим направлениям подготовки / [О. И. Бантикова, В. И. Васянина, Ю. А. Жемчужникова, А.Г. Реннер, Е.Н. Седова, О.И. Стебунова, Л.М. Туктамышева, О.С. Чудинова]; под ред. А. Г. Реннера; Рекомендовано Федеральным государственным автономным образовательным учреждением высшего профессионального образования «Московский физико-технический институт (государственный университет)» в качестве учебного пособия для студентов, обучающихся по программам высшего профессионального образования по экономическим направлениям подготовки. - 2-е изд. - Оренбург : Университет, 2014. - 367 с.
5. Чубукова, И.А. Data Mining / И.А. Чубукова. - 2-е изд., испр. - Москва : Интернет-Университет Информационных Технологий, 2008. - 383 с. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=233055>

Методическая литература

1. Бантикова, О. И. Методы кластерного анализа. Классификация без обучения (непараметрический случай): метод. указания к лаб. практикуму, курсовой работе, диплом. проектированию и самостоят. работе студентов специальности 080116.65, направлений подготовки 231300.62, 080500.62 / О. И. Бантикова, Е. Н. Седова, О. С. Чудинова. - Оренбург: ИПК ГОУ ОГУ, 2011. - 91с.
2. Реннер, А. Г. Параметрический дискриминантный анализ в пакетах Statistica, Stata, Excel: метод. указания к лаб. практикуму, курсовой работе, дипломному проектированию и самостоят. работе студентов / А. Г. Реннер, О. С. Чудинова. - Оренбург: ИПК ГОУ ОГУ, 2011. - 50с.
3. Реннер, А.Г. Снижение размерности признакового пространства методом главных компонент в пакетах Statistica, Stata, Excel: методические указания к лабораторным работам, практическим занятиям и самостоятельной работе студентов / А.Г. Реннер, О.С. Чудинова; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2013. – 46 с.
4. Чудинова, О.С. Расщепление смеси вероятностных распределений в пакете Statistica и среде RStudio [Электронный ресурс]: методические указания / О.С. Чудинова; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2018. – 38 с.

5.3 Периодические издания

1. Вопросы статистики : журнал. - М. : Агентство "Роспечать", 2018.
2. Экономическая наука современной России : информационный бюллетень: журнал. - М.: ВИНТИ, 2018.
3. Маркетинг и маркетинговые исследования : журнал. - М. : Агентство "Роспечать", 2018.

5.4 Интернет-ресурсы

1. www.gks.ru – Федеральная служба государственной статистики;
2. <http://www.machinelearning.ru> – профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных;
3. <https://r-analytics.blogspot.com> – R: Анализ и визуализация данных;
4. <https://yandexdataschool.ru/edu-process/courses/machine-learning> – видеолекции курса «Машинное обучение».
5. <https://www.intuit.ru/studies/courses/1153/318/info> - Национальный Открытый Университет «ИНТУИТ», онлайн-курс «Статистические методы анализа данных» (Автор: Е. Горяинова)
6. <https://www.intuit.ru/studies/courses/11246/1131/info> - Национальный Открытый Университет «ИНТУИТ», онлайн-курс «Обработка экспериментальных данных» (Автор Т. Коваленко)
7. <https://www.intuit.ru/studies/courses/6/6/info> - Национальный Открытый Университет «ИНТУИТ», онлайн-курс «Data Mining» (Автор: И. Чубукова)

8. <https://www.intuit.ru/studies/courses/546/402/info> - Национальный Открытый Университет «ИНТУИТ», онлайн-курс «Прикладная статистика» (Автор: А. Орлов)
9. <https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie> - Образовательная платформа Coursera, онлайн-курс «Введение в машинное обучение»

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы современных информационных технологий

Лицензионное программное обеспечение

1. Операционная система Microsoft Windows
2. Интегрированная система решения математических, инженерно-технических и научных задач PTC MathCAD 14.0 – English

Свободно распространяемое программное обеспечение

1. Open Office/LibreOffice - свободный офисный пакет программ, включающий в себя текстовый и табличный редакторы, редактор презентаций и другие офисные приложения
2. Свободное статистическое программное обеспечение для анализа данных (с открытым исходным кодом в рамках проекта GNU) R
3. Свободная среда разработки программного обеспечения с открытым исходным кодом для языка программирования R Rstudio
4. Прикладной программный пакет для эконометрического моделирования Gretl

Профессиональные базы данных

1. SCOPUS [Электронный ресурс]: реферативная база данных / компания Elsevier. – Режим доступа: <https://www.scopus.com/>, в локальной сети ОГУ
2. Web of Science [Электронный ресурс]: реферативная база данных / компания Clarivate Analytics. – Режим доступа : <http://apps.webofknowledge.com/>, в локальной сети ОГУ

Информационные справочные системы

1. Законодательство России [Электронный ресурс]: информационно-правовая система. – Режим доступа: <http://pravo.fso.gov.ru/ips/>, в локальной сети ОГУ
2. Консультант Плюс [Электронный ресурс] : справочно-правовая система / Компания Консультант Плюс. – Электрон. дан. – Москва, [1992–2018]. – Режим доступа: \\fileserver1\CONSULT\cons.exe, в локальной сети ОГУ
3. Гарант [Электронный ресурс] : справочно-правовая система / НПП Гарант-Сервис. – Электрон. дан. – Москва, [1990–2018]. – Режим доступа: \\fileserver1\GarantClient\garant.exe, в локальной сети ОГУ

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий лекционного типа, семинарского типа, для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Аудитории оснащены комплектами ученической мебели, техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Для проведения лабораторных занятий используется компьютерный класс, оснащенный комплектами ученической мебели, компьютерами с подключением к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду ОГУ.

Помещение для самостоятельной работы обучающихся оснащены компьютерной техникой подключенной к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду ОГУ.

К рабочей программе прилагаются:

- Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине;
- Методические указания для обучающихся по освоению дисциплины.