

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Институт экологии и природопользования



подписано электронно-цифровой подписью

Программа дисциплины
Современные статистические модели М1.В.1

Направление подготовки: 022000.68 - Экология и природопользование

Профиль подготовки: Системная экология и моделирование

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Чижикова Н.А.

Рецензент(ы):

Мухарамова С.С.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Зарипов Ш. Х.

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института экологии и природопользования:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No 224014

Казань
2014

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) доцент, к.н. Чижикова Н.А. кафедра моделирования экологических систем отделение экологии , Nelly.Chizhikova@kpfu.ru

1. Цели освоения дисциплины

формирование представления о современных методах статистического анализа данных в экологии, понимания общих и наиболее важных понятий и математической основы данных методов и развитие практических навыков анализа данных. Все это может быть использовано выпускниками в профессиональной деятельности во всех фундаментальных и практических областях, где требуется анализ наблюдений за состоянием индивидуумов, популяций и сообществ растений и животных, а также требуется оценка взаимосвязи их состояния с природными и антропогенными воздействиями окружающей среды.

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " М1.В.1 Общенаучный" основной образовательной программы 022000.68 Экология и природопользование и относится к вариативной части. Осваивается на 1 курсе, 1 семестр.

Дисциплина "Современные статистические модели" относится к общенаучному блоку М1.В.1 и осваивается на первом году обучения в магистратуре. Она базируется на знаниях и умениях, выработанных при прохождении общих курсов математического и естественнонаучного цикла "Математика", "Информатика", "Теория вероятностей и математическая статистика" бакалавриата, а также курса магистерской программы "Компьютерные технологии и статистические методы в экологии и природопользовании". Полезными для освоения могут являться освоенные курсы профессионального цикла "Введение в количественные методы в экологии", "Классификация наблюдений, экологическая ординация и синтаксономия". Полученные при освоении дисциплины знания и умения могут быть полезны при освоении дисциплин "Технологии географических информационных систем и пространственный анализ", "Современные методы популяционного анализа", "Современные методы экологической индикации состояния окружающей среды". Полученные знания и навыки также могут быть использованы при выполнении научно-исследовательской работы и выпускной работы магистранта.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-4 (общекультурные компетенции)	Владение методами оценки репрезентативности материала, объема выборок при проведении количественных исследований, статистическими методами сравнения полученных данных и определения закономерностей.

В результате освоения дисциплины студент:

1. должен знать:

основы методов анализа данных, применяемых в современной исследовательской практике.

2. должен уметь:

самостоятельно применять

3. должен владеть:

методами и современными программными продуктами анализа данных, применяемыми в современной исследовательской практике.

4. должен демонстрировать способность и готовность:

применять полученные навыки работы с данными, нести ответственность за результаты исследования.

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет зачетных(ые) единиц(ы) 72 часа(ов).

Форма промежуточного контроля дисциплины зачет в 1 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Классификационные и регрессионные деревья (CART).	1	11-12	2	6	0	домашнее задание
2.	Тема 2. Дискриминантный анализ. Многомерный (многофакторный) дисперсионный анализ.	1	13-14	2	6	0	домашнее задание
3.	Тема 3. Методы непараметрической классификации. Иерархическая классификация и метод к-средних.	1	15-16	2	6	0	домашнее задание
	Тема . Итоговая форма контроля	1		0	0	0	зачет
	Итого			6	18	0	

4.2 Содержание дисциплины

Тема 1. Классификационные и регрессионные деревья (CART).

лекционное занятие (2 часа(ов)):

Задачи анализа и моделирования, для которых используется метод классификационных и регрессионных деревьев. Их преимущество перед линейной и нелинейной регрессией. Шкалы измерений зависимых и независимых переменных. Критерии построения правил предсказания в случае регрессионных деревьев (скалярная зависимая переменная), внутригрупповая и межгрупповая дисперсии. Критерии построения правил предсказания в случае классификационных деревьев (номинальная зависимая переменная). Дерево правил. Способы выбора сложности модели. Способы оценки качества полученного дерева решений. Примеры.

практическое занятие (6 часа(ов)):

Анализ данных методом классификационных и регрессионных деревьев с помощью среды статистического программирования R. Функции пакета rpart.

Тема 2. Дискриминантный анализ. Многомерный (многофакторный) дисперсионный анализ.

лекционное занятие (2 часа(ов)):

Задачи анализа и моделирования, для которых используется дискриминантный анализ. Основные предположения об исходных данных. Требования к размеру групп. Дискриминантные функции, интерпретация. Обезразмеривание данных, влияние на результат. Оценка качества разбиения, получаемого с помощью дискриминантного анализа. Задачи анализа и моделирования, для которых используется многофакторный дисперсионный анализ. Основные предположения об исходных данных. Обезразмеривание данных, влияние на результат. Коэффициенты Pillai, Wilks, Hotelling-Lawley. Проверка статистических гипотез многофакторного дисперсионного анализа.

практическое занятие (6 часа(ов)):

Дискриминантный анализ данных с помощью среды статистического программирования R. Функции lda и qda. Многофакторный дисперсионный анализ с помощью среды статистического программирования R. Функция manova.

Тема 3. Методы непараметрической классификации. Иерархическая классификация и метод k-средних.

лекционное занятие (2 часа(ов)):

Задачи анализа, для решения которых используется непараметрическая классификация. Основные предположения об исходных данных. Методы иерархической классификации (дивизивные и аггломеративные), метрики расстояния и сходства (евклидово расстояние, расстояние Махаланобиса, манхеттанское, хордово). Метод Варда. Дендрограммы. Алгоритм метода k-средних. Выбор метода, их плюсы и минусы. Примеры.

практическое занятие (6 часа(ов)):

Иерархическая классификация данных в среде статистического программирования R. Функции hclust, cutree, dist и vegdist. Реализация метода k-средних в среде статистического программирования R. Функция kmeans.

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Классификационные и регрессионные деревья (CART).	1	11-12	подготовка домашнего задания	16	домашнее задание

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
2.	Тема 2. Дискриминантный анализ. Многомерный (многофакторный) дисперсионный анализ.	1	13-14	подготовка домашнего задания	16	домашнее задание
3.	Тема 3. Методы непараметрической классификации. Иерархическая классификация и метод к-средних.	1	15-16	подготовка домашнего задания	16	домашнее задание
	Итого				48	

5. Образовательные технологии, включая интерактивные формы обучения

Разбор учебных примеров (case study) на лекциях, решение учебных задач на компьютере на практических занятиях, выбор методов и решение задач в качестве самостоятельной работы (домашнее творческое задание).

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Классификационные и регрессионные деревья (CART).

домашнее задание , примерные вопросы:

Каждый студент получает на собственных данных классификационное либо регрессионное дерево решений, самостоятельно применяет изученные методы на собственных данных. К сдаче предъявляется написанный студентом код в среде R для реализации анализа, а также иллюстрации полученных результатов в графическом и табличном виде с интерпретацией.

Тема 2. Дискриминантный анализ. Многомерный (многофакторный) дисперсионный анализ.

домашнее задание , примерные вопросы:

Каждый студент проводит самостоятельно дискриминантный и многофакторный дисперсионный анализ, применяет изученные методы на собственных данных. К сдаче предъявляется написанный студентом код в среде R для реализации анализа, а также иллюстрации полученных результатов в графическом и табличном виде с интерпретацией.

Тема 3. Методы непараметрической классификации. Иерархическая классификация и метод к-средних.

домашнее задание , примерные вопросы:

Каждый студент проводит самостоятельно классификацию, применяет изученные методы на собственных данных. К сдаче предъявляется написанный студентом код в среде R для реализации анализа, а также иллюстрации полученных результатов в графическом и табличном виде с интерпретацией.

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

Вопросы на зачет:

Понятие CART.

Различие между регрессионными и классификационными деревьями.
Метод выбора сложности дерева решений.
Задачи дискриминантного анализа.
Необходимые предположения об исходных данных для возможности использования дискриминантного анализа.
Дискриминантная функция, интерпретация.
Влияние обезразмеривания (стандартизации) данных на результат.
Проверка статистических гипотез в рамках многомерного дисперсионного анализа.
Необходимые предположения об исходных данных для возможности использования непараметрической классификации.
Методы непараметрической классификации.
Иерархические методы классификации.
Метрики расстояния и сходства, применяемые для иерархической классификации.
Метод к-средних.
Плюсы и минусы непараметрических методов классификации.

7.1. Основная литература:

1. Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учебное пособие для студентов вузов / В. Е. Гмурман. - 12-е изд., перераб.. - Москва: Юрайт, 2007.
2. Балдин К.В., Рукосуев А.В. Общая теория статистики: учебное пособие. Изд-во: Дашков и К, 2012 г. 312 с. (имеется в библиотеке Знаниум <http://znanium.com/catalog.php?bookinfo=415208>)
3. Основные понятия языка R / Сост. Савельев А.А., Мухарамова С.С., Пилюгин А.Г. Учебно-методическое пособие. - Казань, Изд-во Казан. ун-та. 2007. 30 с. (в свободном доступе в каталоге учебных ресурсов КФУ http://kpfu.ru/publication?p_id=59687)
4. Использование языка R для статистической обработки данных/ Сост. Савельев А.А., Мухарамова С.С., Пилюгин А.Г. Учебно-методическое пособие. - Казань, Изд-во Казан. ун-та. 2007. 30 с. (в свободном доступе в каталоге учебных ресурсов КФУ http://kpfu.ru/publication?p_id=59689)
5. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. 3-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с. - ISBN 978-5-9775-0368-6. (имеется в библиотеке Знаниум <http://znanium.com/bookread.php?book=350638>)
6. Статистика: Учебное пособие / О.А. Шумак, А.В. Гераськин. - М.: ИЦ РИОР: НИЦ Инфра-М, 2012. - 311 с. - ISBN 978-5-369-01048-8 (имеется в библиотеке Знаниум <http://znanium.com/catalog.php?bookinfo=261152>)

7.2. Дополнительная литература:

1. A. Zuur. Analysing ecological data / A. Zuur, E. Ieno, G. Smith. Springer, 2007 (книга полностью доступна из университетской сети по адресу <http://link.springer.com/book/10.1007/978-0-387-45972-1/page/1>).
2. С.А. Айвазян. Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. Москва: Статистика, 1974. 240с.
3. Прикладная статистика. Основы эконометрики: В 2т..?2-е изд., испр.. Учеб. для студ. экон. спец. вузов / С.А.Айвазян, В.С.Мхитарян. М.: ЮНИТИ-ДАНА, 2001.
4. Крамер, Дункан. Математическая обработка данных в социальных науках: современные методы: учеб. пособие для студ. вузов / Д. Крамер. М.: Академия, 2007. - 288 с.

7.3. Интернет-ресурсы:

Дискриминантный анализ с помощью R - <http://www.statmethods.net/advstats/discriminant.html>

Коды, реализующие различные методы анализа из книги A. Zuur "Analysing ecological data" - www.highstat.com/book1.htm

Проверка предположений для проведения классических тестов - <http://www.statmethods.net/stats/anovaAssumptions.html>

Реализация CART моделей с помощью R - <http://www.statmethods.net/advstats/cart.html>

Сайт, посвященный языку и среде статистического программирования R - <http://www.r-project.org/>

Электронная библиотека публикаций - www.springerlink.com

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Современные статистические модели" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, УМК, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

Электронные презентации для проведения занятий.

Программные коды на языке R, реализующие анализ учебных примеров (case studies), разбираемых на занятиях.

Среда статистического программирования R является свободно распространяемым программным приложением и может быть получена каждым студентом с сайта разработчиков <http://cran.r-project.org>.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 022000.68 "Экология и природопользование" и магистерской программе Системная экология и моделирование .

Автор(ы):

Чижикова Н.А. _____

"__" _____ 201__ г.

Рецензент(ы):

Мухарамова С.С. _____

"__" _____ 201__ г.