

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 M1.B.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 2145114

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.
ПК-1 (профессиональные компетенции)	способность формулировать проблемы, задачи и методы научного исследования; получать новые достоверные факты на основе наблюдений, опытов, научного анализа эмпирических данных; реферировать научные труды, составлять аналитические обзоры накопленных сведений в мировой науке и производственной деятельности; обобщать полученные результаты в контексте ранее накопленных в науке знаний; формулировать выводы и практические рекомендации на основе репрезентативных и оригинальных результатах исследований

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-4 (профессиональные компетенции)	использование современных методов обработки и интерпретации экологической информации при проведении научных и производственных исследований

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 продемонстрировать способность и готовность:

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

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 2 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. R как вычислительная среда.	1	11	1	3	0	устный опрос
2.	Тема 2. Линейная модель.	1	12	1	3	0	устный опрос
3.	Тема 3. Нелинейные модели.	1	13	1	3	0	отчет
4.	Тема 4. Методы работы с гетерогенными данными.	1	14	1	3	0	отчет

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
5.	Тема 5. Модели со смешанными эффектами, группы и вложенность.	1	15	1	3	0	отчет
6.	Тема 6. Временная и пространственная корреляция.	1	16	1	3	0	отчет
	Тема . Итоговая форма контроля	1		0	0	0	зачет
	Итого			6	18	0	

4.2 Содержание дисциплины

Тема 1. R как вычислительная среда.

лекционное занятие (1 часа(ов)):

R как вычислительная среда. Структуры данных, чтение и запись таблиц, запись текстовых файлов, создание и сохранение графических иллюстраций. Доступ к справочной системе. Структуры данных - функции `data.frame()`, `list()`, `matrix()`, `array()`, `vector()`, `rep()`, `str()`, `head()`; доступ к элементам. Ввод/вывод - функции `read.table()`, `write.table()`, `hist()`, `plot()`, `points()`, `legend()`. Манипуляции данными ? индексы (логические и числовые), доступ по именам (списки, `dimnames`), обработка данных ? функции `merge()`, `apply()`, `aggregate()`, `tapply()`, `lapply()`, `unlist()`. Выражения, присваивание. Формулы, возможность динамической компиляции.

практическое занятие (3 часа(ов)):

Освоение основных возможностей языка R: структуры данных, чтение и запись таблиц, запись текстовых файлов, создание и сохранение графических иллюстраций. Доступ к справочной системе. Структуры данных - функции `data.frame()`, `list()`, `matrix()`, `array()`, `vector()`, `rep()`, `str()`, `head()`; доступ к элементам. Ввод/вывод - функции `read.table()`, `write.table()`, `hist()`, `plot()`, `points()`, `legend()`. Манипуляции данными ? индексы (логические и числовые), доступ по именам (списки, `dimnames`), обработка данных ? функции `merge()`, `apply()`, `aggregate()`, `tapply()`, `lapply()`, `unlist()`. Выражения, присваивание. Формулы, возможность динамической компиляции.

Тема 2. Линейная модель.

лекционное занятие (1 часа(ов)):

Линейная модель. Факторы и контрасты. Запись моделей (формулы). Функция `lm()` и ее параметры. Анализ результатов, функции доступа к ним ? `summary()`, `coef()`, `residuals()`. Графическое представление результатов и его анализ, функция `plot()`. Предположения и ограничения линейной модели, их проверка (линейность, коллинеарность, гетероскедастичность), обобщенно-линейные модели. Анализ остатков. Выбор параметров, функции `drop1()`, `add1()`. Сравнение моделей, функция `anova()`.

практическое занятие (3 часа(ов)):

Разбор учебного примера на построение и анализ линейной модели. функция `lm()` и ее параметры. Анализ результатов, функции доступа к ним ? `summary()`, `coef()`, `residuals()`. Выбор параметров, функции `drop1()`, `add1()`. Сравнение моделей, функция `anova()`.

Тема 3. Нелинейные модели.

лекционное занятие (1 часа(ов)):

Нелинейные модели. Библиотеки `mgcv` и `gam`. Непараметрическая модель `loess()`. Сплайны, сплайны с штрафом, перекрестная проверка и обобщенная перекрестная проверка.

практическое занятие (3 часа(ов)):

Разбор учебного примера на построение нелинейной модели. Работа с функцией `gam()`. Непараметрическая модель `loess()`. Сплаины, сплайны с пенальти, перекрестная проверка и обобщенная перекрестная проверка.

Тема 4. Методы работы с гетерогенными данными.

лекционное занятие (1 часа(ов)):

Методы работы с гетерогенными данными. Модели дисперсии `varIdent()`, `varPower()`, `varExp()`, `varConstPower()`, `varComb()`.

практическое занятие (3 часа(ов)):

Разбор учебного примера для создания модели с учетом гетероскедастичности. Модели дисперсии `varIdent()`, `varPower()`, `varExp()`, `varConstPower()`, `varComb()`.

Тема 5. Модели со смешанными эффектами, группы и вложенность.

лекционное занятие (1 часа(ов)):

Модели со смешанными эффектами, группы и вложенность. Линейные модели со смешанными эффектами, функция `glm()`. Модель со случайной константой. Модель со случайным наклоном. Индуцированная корреляция. Оценки максимального правдоподобия и ограниченного максимального правдоподобия. Протокол выбора оптимальной модели. Валидация модели.

практическое занятие (3 часа(ов)):

Разбор учебного примера для создания модели со смешанными эффектами. Функция `glm()`. Валидация модели.

Тема 6. Временная и пространственная корреляция.

лекционное занятие (1 часа(ов)):

Временная и пространственная корреляция. Стандартная временная корреляция - модель ARMA. Пространственная корреляция и ее потенциальные причины. Анализ и моделирование временной и пространственной корреляции.

практическое занятие (3 часа(ов)):

Разбор учебного примера для создания модели временной корреляцией с использованием функций `ar()`, `arma()`.

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. R как вычислительная среда.	1	11	подготовка домашнего задания	0	домашнее задание
				подготовка к устному опросу	8	устный опрос
2.	Тема 2. Линейная модель.	1	12	подготовка домашнего задания	0	домашнее задание
				подготовка к устному опросу	8	устный опрос
3.	Тема 3. Нелинейные модели.	1	13	подготовка домашнего задания	0	домашнее задание
				подготовка к отчету	8	отчет

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
4.	Тема 4. Методы работы с гетерогенными данными.	1	14	подготовка к отчету	8	отчет
5.	Тема 5. Модели со смешанными эффектами, группы и вложенность.	1	15	подготовка к отчету	8	отчет
6.	Тема 6. Временная и пространственная корреляция.	1	16	подготовка к отчету	8	отчет
	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. R как вычислительная среда.

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.

устный опрос , примерные вопросы:

Примерный вопрос: В чем основное различие массива данных (data frame) и матрицы (matrix)?

Тема 2. Линейная модель.

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.

устный опрос , примерные вопросы:

Примерный вопрос: Какие четыре предположения необходимо проверить для того, чтобы строить линейную модель?

Тема 3. Нелинейные модели.

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.

отчет , примерные вопросы:

Студенту дается учебный пример (набор данных) для построения нелинейной модели. По выполненному заданию пишется отчет с описанием команд и результатов.

Тема 4. Методы работы с гетерогенными данными.

отчет , примерные вопросы:

Студенту дается учебный пример (набор данных) для построения модели, учитывающей гетероскедастичность данных. Студенту необходимо выбрать модель дисперсии. По выполненному заданию пишется отчет с описанием команд и результатов.

Тема 5. Модели со смешанными эффектами, группы и вложенность.

отчет , примерные вопросы:

Студенту дается учебный пример (набор данных) для анализа смешанных эффектов. Студенту необходимо выбрать модель смешанных эффектов. По выполненному заданию пишется отчет с описанием команд и результатов.

Тема 6. Временная и пространственная корреляция.

отчет , примерные вопросы:

Студенту дается учебный пример (набор данных) для построения модели с учетом временной корреляции. Студенту необходимо выбрать порядок (лаг), на котором наблюдается автокорреляция. По выполненному заданию пишется отчет с описанием команд и результатов.

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

Примерные вопросы для зачета:

Линейная модель, реализация в среде R.

Факторы и контрасты.

Анализ построенной линейной модели, возможности среды R.

Предположения и ограничения линейной модели, их проверка (линейность, коллинеарность, гетероскедастичность), обобщенно-линейные модели.

Анализ остатков для оценки качества линейной модели.

Выбор предикторов в линейной модели, функции `drop1()`, `add1()` в среде R.

Сравнение моделей, функция `anova()`.

Нелинейные модели.

Библиотеки `mgcv` и `gam`.

Непараметрическая модель `loess()`.

Сплайны, сплайны с пенальти, перекрестная проверка и обобщенная перекрестная проверка при выборе сложности нелинейной модели.

Методы работы с гетерогенными данными.

Модели дисперсии `varIdent()`, `varPower()`, `varExp()`, `varConstPower()`, `varComb()`.

Модели со смешанными эффектами, группы и вложенность.

Линейные модели со смешанными эффектами, функция `glm()`.

Модель со случайной константой.

Модель со случайным наклоном.

Индукцированная корреляция.

Оценки максимального правдоподобия и ограниченного максимального правдоподобия.

Протокол выбора оптимальной модели.

Валидация модели.

Временная и пространственная корреляция.

Стандартная временная корреляция - модель ARMA.

Пространственная корреляция и ее потенциальные причины.

Анализ и моделирование временной и пространственной корреляции.

7.1. Основная литература:

Наглядная статистика. Используем R!, Шипунов, Алексей Борисович;Балдин, Евгений Михайлович;Волкова, Полина Андреевна, 2012г.

2. Гусева, Е. Н. Теория вероятностей и математическая статистика [Электронный ресурс] : Уч. пособ. / Е. Н. Гусева. - 5-е изд., стереотип. - М. : Флинта, 2011. - 220 с.

<http://znanium.com/bookread.php?book=406064>

3. 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>)

4. 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>)

5. A. Zuur. Mixed Effects Models and Extensions in Ecology with R / A. Zuur, E. Ieno, N. Walker, A. Saveliev. Springer, 2009 (книга полностью доступна по подписке университета в электронной библиотеке издательства Springer <http://link.springer.com/book/10.1007/978-0-387-87458-6>).

7.2. Дополнительная литература:

Оптимальные статистические решения, Володин, Игорь Николаевич, 2012г.

2. Буре В. М., Парилина Е. М. Теория вероятностей и математическая статистика. Лань, 2013. 416 с. (http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=10249)

3. Савельев А.А., Мухарамова С.С., Пилюгин А.Г., Чижикова Н.А. Геоэкономический анализ данных в экологии и природопользовании (с применением пакета R). - Казань: Казанский ун-т, 2012. - 120с. http://kpfu.ru/publication?p_id=59674,
<http://shelly.ksu.ru/e-ksu/docs/F1387902447/geostat-2012.pdf>

7.3. Интернет-ресурсы:

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

02 Коды, реализующие различные методы анализа из книги A. Zuur - www.highstat.com/book1.htm

03 Quick R - кратко о некоторых статистических методах - <http://www.statmethods.net/>

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

05 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 "БиблиоРоссика", доступ к которой предоставлен студентам. В ЭБС "БиблиоРоссика" представлены коллекции актуальной научной и учебной литературы по гуманитарным наукам, включающие в себя публикации ведущих российских издательств гуманитарной литературы, издания на английском языке ведущих американских и европейских издательств, а также редкие и малотиражные издания российских региональных вузов. ЭБС "БиблиоРоссика" обеспечивает широкий законный доступ к необходимым для образовательного процесса изданиям с использованием инновационных технологий и соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

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.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен студентам. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

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.