

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



подписано электронно-цифровой подписью

Программа дисциплины

Статистическая обработка и анализ данных

Направление подготовки: 19.03.01 - Биотехнология

Профиль подготовки: Биотехнология и биоинженерия

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): доцент, к.н. Никоненкова Т.В. (Кафедра моделирования экологических систем, Отделение экологии), Tatjana.Nikonenkova@kpfu.ru ; доцент, к.н. (доцент) Чижикова Н.А. (Кафедра моделирования экологических систем, Отделение экологии), Nelly.Chizhikova@stud.kpfu.ru ; Пилогин Александр Геннадиевич

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОПК-2	Способен осуществлять поиск, хранение, обработку и анализ профессиональной информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий, включая проведение расчетов и моделирование, с учетом основных требований информационной безопасности
ОПК-3	Способен принимать участие в разработке алгоритмов и программ, пригодных для практического применения в сфере своей профессиональной деятельности
ОПК-7	Способен проводить экспериментальные исследования и испытания по заданной методике, наблюдения и измерения, обрабатывать и интерпретировать экспериментальные данные, применяя математические, физические, физико-химические, химические, биологические, микробиологические методы

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

основы языка R для статистической обработки данных и работы с графикой, основные алгоритмы обработки данных и способы их внешнего представления

Должен уметь:

ориентироваться в основных алгоритмах статистической обработки данных для решения практических задач и графического представления результатов

Должен владеть:

приемами программирования, компьютерными методами обработки данных экологических исследований и мониторинга

Должен демонстрировать способность и готовность:

применять полученные знания и навыки на практике, нести ответственность за результат выполнения работы.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.О.32 Дисциплины (модули)" основной профессиональной образовательной программы 19.03.01 "Биотехнология (Биотехнология и биоинженерия)" и относится к обязательной части ОПОП ВО.

Осваивается на 3 курсе в 5 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) на 108 часа(ов).

Контактная работа - 63 часа(ов), в том числе лекции - 18 часа(ов), практические занятия - 44 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 1 часа(ов).

Самостоятельная работа - 27 часа(ов).

Контроль (зачёт / экзамен) - 18 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 5 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Се-местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само-стоя-тель-ная ра-бота
			Лекции, всего	Лекции в эл. форме	Практи-ческие занятия, всего	Практи-ческие в эл. форме	Лабора-торные работы, всего	Лабора-торные в эл. форме	
1.	Тема 1. Введение в R	5	2	0	6	0	0	0	3
2.	Тема 2. Описательная статистика и статистические критерии в R	5	8	0	20	0	0	0	12
3.	Тема 3. Корреляционный и регрессионный анализ в R	5	4	0	10	0	0	0	6
4.	Тема 4. Классификация и кластеризация в R	5	4	0	8	0	0	0	6
	Итого		18	0	44	0	0	0	27

4.2 Содержание дисциплины (модуля)

Тема 1. Введение в R

Лекция 1. Язык программирования R. Установка Rgui и RStudio. Интерфейс RStudio. Типы данных: Векторы и матрицы, Факторы, Списки и таблицы. Типы шкал, в которых могут быть представлены данные. Преобразование разных типов и структур данных друг в друга. Индексация. Арифметические и логические операции, формулы. Пропущенные наблюдения. Импортирование данных в R.

Тема 2. Описательная статистика и статистические критерии в R

Лекция 2. Генеральная совокупность и выборка. Точечная оценка выборочных параметров распределения. Меры центральной тенденции (выборочное среднее, мода, медиана). Меры изменчивости (размах, дисперсия, среднее квадратическое отклонение, коэффициент вариации). Меры формы (асимметрия, эксцесс). Квантиль, квартиль распределения. Доверительные интервалы для математического ожидания и дисперсии. Графический анализ данных: диаграмма рассеяния, гистограмма, Box Plot. Статистическая гипотеза. Статистические критерии. P-value. Алгоритм проверки статистических гипотез.

Лекция 3. Законы распределения вероятностей. Статистические критерии проверки на нормальность распределения. Графические способы проверки на нормальность распределения.

Лекция 4. Статистические критерии: параметрические и непараметрические критерии. Критерии для сравнения дисперсий двух нормальных генеральных совокупностей. Критерии сравнения средних двух нормальных генеральных совокупностей. Критерии для сравнения дисперсий нескольких нормальных генеральных совокупностей. Критерии сравнения средних нескольких нормальных генеральных совокупностей. Критерии для проверки гипотез о равенстве дисперсий/средних генеральных совокупностей, которые не подчиняются нормальному закону распределения.

Лекция 5. Анализ категориальных данных в R. Представление категориальных данных в виде таблиц частот и таблиц сопряженности. Визуализация категориальных данных. Критерии для проверки независимости категориальных данных. Работа с отсутствующими значениями (NA) в R. Анализ выбросов в R.

Тема 3. Корреляционный и регрессионный анализ в R

Лекция 6. Визуализация данных в R. Базовые графики в R с функцией plot(). Основы ggplot. Условные оператор if, ifelse. Циклы for и while. Циклы на основе apply. Устройство функций в R. Написание пользовательских функций в R. Агрегатные функции в R.

Лекция 7. Коэффициенты корреляции Пирсона, Спирмена, Кендела в R. Корреляционные матрицы в R. Визуализация корреляций между переменными средствами R. Простая линейная регрессия в R. Множественная регрессия в R. Оценка качества линейной регрессионной модели. Ограничения линейной регрессии. Прогнозирование с помощью линейной регрессии.

Тема 4. Классификация и кластеризация в R

Лекция 8. Классическое машинное обучение. Машинное обучение с учителем и машинное обучение без учителя. Метрический классификатор. Метрики в пространстве признаков (мера близости объектов). Алгоритм k ближайших соседей.

Лекция 9. Кластерный анализ. Метод k-средних. Иерархический кластерный анализ. Методы иерархической кластеризации (методы ближайшего соседа, дальнего соседа, центроидов, метод Варда, ...). Нормирование данных. Предварительная обработка наблюдений. Дендрограммы.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемому результату обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. R: анализ и визуализация данных - <https://r-analytics.blogspot.com/>
2. Портал о машинном обучении - <http://www.machinelearning.ru/>
4. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R - <https://r-analytics.blogspot.com/2014/12/r.html>
4. Национальная электронная библиотека - <https://rusneb.ru/>
5. КонсультантПлюс - <http://www.consultant.ru/cons/cgi/online.cgi>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	перед очередной лекцией необходимо просмотреть по конспекту материал предыдущей лекции. При затруднениях в восприятии материала надо обратиться к основным литературным источникам. Если разобраться в материале опять не удалось, обратиться к лектору по графику его консультаций или на практических занятиях; - хотя бы бегло ознакомиться с содержанием очередной лекции по основным источникам литературы в соответствии с рабочей программой дисциплины.
практические занятия	- на занятия носить рекомендованную лектором литературу; - до очередного практического занятия по конспекту (или литературе) проработать теоретический материал, соответствующий темы занятия; - в начале занятий задать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и освоении при решении задач, заданных для самостоятельного решения; - иметь при себе конспект лекций
самостоятельная работа	- выполнять все плановые задания, выдаваемые преподавателем для самостоятельного выполнения, и разбирать на практических занятиях и консультациях неясные вопросы; - выполнять и сдавать в срок письменную домашнюю работу; - подготовку к экзамену необходимо проводить по экзаменационным теоретическим вопросам; - при подготовке к экзамену параллельно прорабатываете соответствующие теоретические и практические разделы курса, все неясные моменты фиксируйте и выносите на плановую консультацию
экзамен	При подготовке к итоговому контролю изучить основную и дополнительную литературу, изучить конспекты лекций и материал наработанный во время практических занятий, устного опроса и выполнения домашнего задания, проработать менее запомнившиеся темы и вопросы, которые возникали во время изучения дисциплины. Вынести вопросы на плановую консультацию

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Мультимедийная аудитория.

Компьютерный класс.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 19.03.01 "Биотехнология" и профилю подготовки "Биотехнология и биоинженерия".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.О.32 Статистическая обработка и анализ данных

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 19.03.01 - Биотехнология

Профиль подготовки: Биотехнология и биоинженерия

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Основная литература:

1. Статистические методы анализа данных : учебник / Л.И. Ниворожкина, С.В. Арженовский, А.А. Рудяга [и др.] ; под общ. ред. д-ра экон. наук, проф. Л.И. Ниворожкиной. - Москва: РИОР : ИНФРА-М, 2016. - 333 с. - (Высшее образование: Бакалавриат). - www.dx.doi.org/10.12737/21064. - ISBN 978-5-369-01612-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/556760> (дата обращения: 05.03.2020). - Режим доступа: по подписке.
2. Козлов, А.Ю. Статистический анализ данных в MS Excel : учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. - Москва: ИНФРА-М, 2017. - 320 с. - (Высшее образование: Бакалавриат). - www.dx.doi.org/10.12737/2842. - ISBN 978-5-16-101024-2. - Текст : электронный. - URL: <http://znanium.com/catalog/product/858510> (дата обращения: 05.03.2020). - Режим доступа: по подписке.
3. Статистический анализ данных в экологии и природопользовании с использованием программы STATGRAPHICS Plus: учебно-методическое пособие / Казан. (Приволж.) федер. ун-т, Фак. географии и экологии ; сост.: К. А. Мальцев, С. С. Мухарамова. - Казань : [Казанский университет], 2011. - 34 с. - Текст : электронный. - URL : http://kpfu.ru/docs/F1860770149/mss1_stat_Uchebno_metodich_posobie.pdf (дата обращения: 05.03.2020). - Режим доступа: открытый.

Дополнительная литература:

1. Ефимова, М. Р. Общая теория статистики: учебник / М.Р. Ефимова, Е.В. Петрова, В.Н. Румянцев. - 2-е изд., испр. и доп. - Москва : ИНФРА-М, 2011. - 416 с. (Высшее образование). ISBN 978-5-16-004265-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/251320> (дата обращения: 27.02.2020). - Режим доступа: по подписке.
2. Гвоздева, В. А. Базовые и прикладные информационные технологии: учебник / В. А. Гвоздева. - Москва: ФОРУМ: ИНФРА-М, 2020. - 384 с. - (Высшее образование). - ISBN 978-5-8199-0572-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1053944> (дата обращения: 27.02.2020). - Режим доступа: по подписке.
3. Кулаичев, А. П. Методы и средства комплексного статистического анализа данных : учебное пособие / А.П. Кулаичев. - 5-е изд., перераб. и доп. - Москва : ИНФРА-М, 2018. - 484 с. - (Высшее образование: Бакалавриат). - www.dx.doi.org/10.12737/25093. - ISBN 978-5-16-012834-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/975598> (дата обращения: 05.03.2020). - Режим доступа: по подписке.
4. Мельниченко, А. С. Математическая статистика и анализ данных: учебное пособие / А. С. Мельниченко. - Москва : МИСИС, 2018. - 45 с. - ISBN 978-5-906953-62-9. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/108035> (дата обращения: 05.03.2020). - Режим доступа: для авториз. пользователей.
5. Основные понятия языка R : учебно-методическое пособие / Казан. гос. ун-т, Фак. географии и экологии ; сост. А. А. Савельев, С. С. Мухарамова, А. Г. Пилюгин, Е. А. Алексеева. - Казань : Казанский государственный университет, 2007. - 29 с. - Текст : электронный. - URL: http://kpfu.ru/docs/F568269105/metodichka_R_1.pdf (дата обращения: 05.03.2020). - Режим доступа: открытый.

*Приложение 3
к рабочей программе дисциплины (модуля)
Б1.О.32 Статистическая обработка и анализ данных*

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 19.03.01 - Биотехнология

Профиль подготовки: Биотехнология и биоинженерия

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. 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, учебно-методические комплексы, 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 образования (ФГОС ВО) нового поколения.