

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»



УТВЕРЖДЕНО

Первый проректор

А.В. Толстикова А.В. Толстикова

29 марта 2022 г.

**Компьютерные технологии и статистические методы
в геоэкологических исследованиях**
Рабочая программа
для обучающихся по научной специальности
1.6.21. Геоэкология
форма обучения (очная)

Осипова Н.Г., Синдирева А.В. **Компьютерные технологии и статистические методы в геоэкологических исследованиях.** Рабочая программа для обучающихся по научной специальности 1.6.21. Геоэкология, форма обучения (очная). Тюмень, 2022.

Рабочая программа составлена в соответствии с требованиями ФГТ, утвержденными приказом Минобрнауки от 20.10.2021 № 951.

Рабочая программа **Компьютерные технологии и статистические методы в геоэкологических исследованиях** опубликована на сайте ТюмГУ: [электронный ресурс] / Режим доступа: <https://www.utmn.ru/sveden/education/#>.

1. Цели и задачи освоения дисциплины

Цель дисциплины - внедрение в образовательный процесс инновационных компьютерных технологий и современного программного обеспечения для подготовки высококвалифицированных кадров в экологии и природопользовании.

Задачи дисциплины:

- изучение современной методологии статистического анализа и компьютерного моделирования для получения теоретических зависимостей на основе экспериментальных данных в экологии и природопользовании;
- изучение дисперсионного, кластерного, корреляционного и регрессионного анализа, анализа временных рядов экспериментальных данных;
- формирование навыков работы в языке программирования R;
- проведение статистического анализа с применением соответствующих методов, моделирование, интерпретация полученных результатов; создание прогнозов в R.

2. Компетенции обучающегося, формируемые в результате освоения данной дисциплины

- **ПК-17** - знать современные междисциплинарные проблемы геоэкологии и использовать фундаментальные представления о составе, строении, свойствах, процессах, структуре и функционировании геосфер Земли как среды обитания человека и других организмов в сфере профессиональной деятельности;
- **ПК-18** - уметь использовать основные теории, концепции и принципы в области геоэкологической деятельности, быть способным к системному мышлению, демонстрировать знание истории и методологии геоэкологии, расширяющие общепрофессиональную, фундаментальную подготовку;
- **ПК-19** - владеть методами, самостоятельно анализировать необходимую информацию с использованием современных информационных технологий, выявлять фундаментальные проблемы и выполнять разномасштабные научные и прикладные исследования в профессиональной области.

3. Структура и объем дисциплины

Таблица 1

Вид учебной работы	Всего часов (академические часы)	Часов в семестре (академические часы)
		4
Общий объем	зач. ед.	3
	час	108
Из них:		

Часы аудиторной работы (всего):	22	22
Лекции	12	12
Практические занятия	10	10
Лабораторные / практические занятия по подгруппам		
Часы внеаудиторной работы, включая самостоятельную работу обучающегося	50	50
Вид промежуточной аттестации (диф. зачет, кандидатский экзамен, экзамен)	36	Дифференцированный зачет 36

4. Система оценивания

4.1. Необходимым условием допуска к дифференцированному зачёту является посещение лекций и сдача всех практических занятий по дисциплине. Аспиранты сдают дифференцированный зачёт в устной форме.

Критерии оценки ответов:

- **оценка «отлично»** выставляется обучающемуся, который дает полный, развернутый ответ на поставленный вопрос, показывает совокупность осознанных знаний о статистических методах анализа данных об объекте или явлении, может привести пример. Свободно владеет компьютерными технологиями для анализа данных, создает модели, правильно интерпретирует результаты моделирования, умеет делать прогнозы.

- **оценка «хорошо»** выставляется обучающемуся, который дает полный, развернутый ответ на поставленный вопрос, показывает совокупность осознанных знаний о статистических методах анализа данных об объекте или явлении, может привести пример. Владеет компьютерными технологиями для анализа данных, создает модели, правильно интерпретирует результаты моделирования, умеет делать прогнозы, но допускает не большие неточности.

- **оценка «удовлетворительно»** выставляется обучающемуся, который дает ответ на поставленный вопрос, имеет знания о статистических методах анализа данных об объекте или явлении. Владеет компьютерными технологиями для анализа данных, создает модели, правильно интерпретирует результаты моделирования, умеет делать прогнозы, но допускает неточности

- **оценка «неудовлетворительно»** выставляется обучающемуся за отсутствие ответа.

5. Содержание дисциплины

5.1. Тематический план дисциплины

Таблица 2

№ п/ п	Наименование тем и/или разделов	Объем дисциплины (модуля), час.				
		Всего	Виды аудиторной работы (академические часы)			Иные виды контактн ой работы
			Лекци и	Практичес кие занятия	Лабораторн ые/ практически е занятия по подгруппам	
1	2	3	4	5	6	7
1.	Введение в анализ данных	4	2	2	0	0
2.	Проверка статистических гипотез	4	2	2	0	0
3	Корреляционно - регрессионный анализ в экологическом мониторинге	4	2	2	0	0
4	Обзор ПО для анализа данных. R и статистика	4	2	2	0	0
5	Анализ экологических данных в R	4	2	2	0	0
6	Дисперсионный анализ экологических данных	2	2	0	0	0
7	Дифференцирован ный зачет	36	0	0	0	36
	Итого (часов)	58	12	10	0	36

5.2. Содержание дисциплины (модуля) по темам

Тема 1. Введение в анализ данных

Основные понятия методов обработки и анализа данных, базирующиеся на положениях теории вероятностей, математической статистики и теории случайных процессов. Предмет, метод, задачи статистики в экологическом мониторинге.

Статистическое наблюдение. Группировка и представление данных. Статистические показатели. Анализ вариационного и динамического рядов распределения.

Учебно-исследовательская работа

Представление статистических данных, построение вариационных рядов, вычисление средних величин и показателей вариации.

Типы статистических данных. Генеральная совокупность и выборка. Оценка характеристик генеральной совокупности по выборке. Точечные и интервальные оценки. Доверительные интервалы. Описательная характеристика столбцов матрицы экспериментов. Проверка гипотез о нормальности и равномерности распределения случайной величины.

Тема 2. Проверка статистических гипотез

Выборочное оценивание. Несмещенные, эффективные и состоятельные оценки. Интервальное оценивание. Доверительные интервалы для выборочного среднего и выборочной дисперсии. Распределения Стьюдента и "хи-квадрат". Ошибки первого и второго рода. Гипотеза о равенстве двух выборочных средних, двух выборочных дисперсий. Критерии Пирсона, Колмогорова.

Мастерская

Ключевые слова. Работа с векторами. Функции: `sort`, `length`, `rev`, `min`, `max`, `sum`, `mean` и др.

Работа с матрицами: функции `rowSums`, `rowMeans`, `colSums`, `colMeans`; `det` - определитель, `ginv` - обратная матрица, `cbind` - добавление столбца, `rbind` - добавление строки, `%*%` - умножение матриц. Работа с многомерными массивами.

Работа с факторами/признаками, с наборами элементов разных типов. Таблицы данных: `Data frame`. Создание из файла: `dt<-read.table("имя файла")`;

- из векторов: `dt<-data.frame(v1,v2)`;

Работа: `dt`; `attach`, `summary`. Примеры

Тема 3. Корреляционно - регрессионный анализ в экологическом мониторинге

Корреляционная зависимость. Выборочный коэффициент корреляции, методика его вычисления. Корреляционная связь. Значимость коэффициента корреляции. Линейная регрессия. Подбор параметров выборочного уравнения прямой линии регрессии. Метод наименьших квадратов. Нелинейная регрессия. Множественная регрессия.

Мастерская

Анализ временных рядов

Построение временных рядов. Основные характеристики и компоненты временного ряда. Показатели временного ряда и их вычисление. Определение тренда и сглаживания временного ряда. Экспоненциальное сглаживание и прогнозирование.

Тема 4. Обзор ПО для анализа данных. R и статистика.

Обзор ППП для обработки и анализа данных: `Excel`, `Statistica`, `SPSS`, `R` и др. Обзор систем компьютерного моделирования: `Matlab`, `Arena`, `GPSS`, `R` и др. Достоинства и обоснование выбора `R` для обработки данных в экологическом мониторинге. Интерфейс и функциональные возможности языка программирования `R`. Типы данных в `R`.

Основные операторы языка программирования `R`.

Тема 5. Анализ экологических данных в R.

Анализ структуры данных. Многомерные данные. Дискриминантный анализ. Множественная регрессия в `R`. Моделирование экологических задач, оценка моделей, принятие решений, прогнозирование.

Анализ временных рядов и прогнозирование. Виды и построение временных рядов. Основные характеристики и компоненты временного ряда. Показатели временного ряда и их вычисление. Определение тренда и сглаживания временного ряда. Экспоненциальное сглаживание и прогнозирование.

Тема 6. Дисперсионный анализ экологических данных.

Дисперсионный анализ. Примеры экологических задач с применением дисперсионного анализа и их реализация в R. Критерий Стьюдента. Критерий Уилкоксона.

6. Учебно-методическое обеспечение и планирование самостоятельной работы обучающихся

Таблица 3

№ Темы	Темы	Формы СРС, включая требования к подготовке к занятиям
1.	Основы статистики	Чтение обязательной и дополнительной литературы
2.	Проверка статистических гипотез	Чтение обязательной и дополнительной литературы, решение задач
3.	Проверка статистических гипотез	Проработка лекций
4.	Факторный анализ	Чтение обязательной и дополнительной литературы
5.	Корреляционно - регрессионный анализ	Чтение обязательной и дополнительной литературы
6.	Корреляционно-регрессионный анализ	Проработка лекций
7.	Корреляционно - регрессионный анализ	Проработка лекций
8.	Анализ временных рядов и прогнозирование	Чтение обязательной и дополнительной литературы
9.	Обзор ПО для анализа данных	Чтение обязательной и дополнительной литературы
10.	Интерфейс RStudio	Проработка лекций
11.	R и статистика	Чтение обязательной и дополнительной литературы
12.	Основы синтаксиса языка R. Типы данных.	Проработка лекций, выполнение практических заданий
13.	Таблицы данных в R	Проработка лекций
14.	Статистика данных в R	Проработка лекций, выполнение практических заданий
15.	Операторы ветвления. Циклы	Проработка лекций, выполнение практических заданий

16.	Библиотека dplyr	Проработка лекций, выполнение практических заданий
17.	Корреляционно-регрессионный анализ в R	Проработка лекций, выполнение практических заданий
18.	Оценка регрессионной модели	Проработка лекций, выполнение практических заданий
19.	Нелинейная регрессионная модель	Проработка лекций, выполнение практических заданий
20.	Прогнозирование в R	Проработка лекций, выполнение практических заданий
21.	Дисперсионный анализ в R	Проработка лекций, выполнение практических заданий
22.	Базовая графика в R	Проработка лекций, выполнение практических заданий
23.	Проверка знаний	Проработка лекций, выполнение практических заданий
24.	Дифференцированный зачет	Ответы на вопросы

7. Промежуточная аттестация по дисциплине (модулю)

7.1 Оценочные материалы для проведения промежуточной аттестации по дисциплине (модулю)

Вопросы к дифференцированному зачету

1. Классификация случайных событий. Классическое определение вероятности. Свойства вероятности события. Примеры.
2. Несовместные и совместные события. Сумма событий. Теоремы сложения вероятностей.
3. Зависимые и независимые события. Произведение событий. Понятие условной вероятности. Теорема умножения вероятностей.
4. Формула полной вероятности. Формула Байеса.
5. Повторные независимые испытания. Формула Бернулли. Наивероятнейшее число наступлений события.
6. Понятие случайной величины и ее описание. Дискретная случайная величина и закон ее распределения. Независимые случайные величины.
7. Математическое ожидание дискретной случайной величины и его свойства. Дисперсия дискретной случайной величины и ее свойства.

8. Функция распределения случайной величины, ее определение, свойства и график.
9. Непрерывная случайная величина (НСВ). Вероятность отдельно взятого значения НСВ. Плотность вероятности НСВ. Математическое ожидание и дисперсия НСВ.
10. Случайная величина, распределенная по биномиальному закону, ее математическое ожидание и дисперсия. Геометрическое распределение.
11. Закон распределения Пуассона. Математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона.
12. Равномерный закон распределения.
13. Показательный (экспоненциальный) закон распределения.
14. Нормальный закон распределения. Функция распределения нормально распределенной случайной величины и ее выражение через функцию Лапласа.
15. Формулы для определения вероятности: а) попадания нормально распределенной случайной величины в заданный интервал; б) ее отклонения от математического ожидания. «Правило трех сигм».
16. Закон больших чисел. Теорема Бернулли и ее значение. Центральная предельная теорема.
17. Основы статистики. Проверка статистических гипотез. Выборочное оценивание. Несмещенные, эффективные и состоятельные оценки. Интервальное оценивание
18. Доверительные интервалы для выборочного среднего и выборочной дисперсии. Распределения Стьюдента и "хи-квадрат". Ошибки первого и второго рода.
19. Среда R: сведения из истории создания, источники получения, меню и структура пакета, настройка пользовательской конфигурации, редактор данных, визуализация данных.
20. Типы данных в среде R. Основные операторы.
21. Предварительная обработка данных в задачах анализа: нормализация, стандартизация, обработка пропущенных значений.
22. Способы получения репрезентативных выборок. Реализация в среде R.
23. Корреляционный анализ. Коэффициент корреляции. Критерий Стьюдента. Построение доверительного интервала.
24. Линейный регрессионный анализ: постановка задачи, метод наименьших квадратов, свойства решения, приложения. Реализация в среде R.
25. Однофакторная линейная регрессионная модель: графическая интерпретация решения задачи, исследование признаков на независимость. Реализация в среде R.
26. Множественная линейная регрессия: постановка и решение задачи, примеры применения модели. Несмещённость оценок. Реализация в среде R.
27. Задачи классификации: общая постановка, виды, обзор методов решения, возможные приложения.

28. Метод k-ближайших соседей: постановка задачи, идея и алгоритм метода, графическая иллюстрация, вычислительные аспекты, примеры применения. Реализация в среде R.
29. Понятие деревьев решений и смежные понятия. Решение задачи классификации методом построения деревьев решений. Реализация в среде R.
30. Линейный дискриминантный анализ: постановка задачи, идея и алгоритм метода, графическая иллюстрация (для однофакторной задачи), примеры применения. Реализация в среде R.

8. Учебно-методическое и информационное обеспечение дисциплины (модуля)

8.1 Основная литература:

1. Трифонова, Т. А. Геоинформационные системы и дистанционное зондирование в экологических исследованиях: учебное пособие для вузов / Т. А. Трифонова, Н. В. Мищенко, А. Н. Краснощеков. — М.: Академический Проект, 2015. — 350 с. — ISBN 978-5-8291-0602-7. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <http://www.iprbookshop.ru/60288.html>. Дата обращения: 22.03.2022

8.2 Дополнительная литература:

1. Волков А.В. Географические информационные системы [Электронный ресурс] : учебное пособие / А.В. Волков, М.М. Орехов. — Электрон. текстовые данные. — Санкт-Петербург: Санкт-Петербургский государственный архитектурно-строительный университет, ЭБС АСВ, 2015. — 76 с. — 978-5-9227-0600-1. — Режим доступа: <http://www.iprbookshop.ru/58532.html> Дата обращения: 22.03.2022

2. Системный анализ и математическое моделирование сложных экологических и экономических систем. Теоретические основы и приложения : монография / О. Е. Архипова, В. Ю. Запорожец, О. В. Ковалев [и др.] ; под ред. Ф. А. Сурков, В. В. Селютин. — Ростов-на-Дону: Издательство Южного федерального университета, 2015. — 162 с. — ISBN 978-5-9275-1985-9. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <http://www.iprbookshop.ru/78703.html>. Дата обращения: 22.03.2022

3. Математическое и компьютерное моделирование в экологии: учебное пособие / С. В. Бобырев, А. В. Косарев, А. Л. Подольский [и др.]. — Саратов: Саратовский государственный технический университет имени Ю.А. Гагарина, ЭБС АСВ, 2012. — 106 с. — ISBN 2227-8397. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <http://www.iprbookshop.ru/76487.html> Дата обращения: 22.03.2022

8.3 Интернет-ресурсы: (при необходимости)

<https://www.esri-cis.ru/ru-ru/home>

<http://www.esri.com>

<https://learn.arcgis.com/ru/gallery/>

9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) Лицензионное ПО;

ПО, находящееся в свободном доступе: MS Office, платформа для электронного обучения Microsoft Teams, R-Studio, язык программирования R.

10. Технические средства и материально-техническое обеспечение дисциплины (модуля)

– *Учебные аудитории для проведения лекций и практических занятий;*

11. Средства адаптации преподавания дисциплины (модуля) к потребностям лиц с ограниченными возможностями

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться одни из следующих вариантов восприятия информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); индивидуальные задания и консультации.

- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные задания и консультации.

- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.

12. Методические рекомендации обучающимся по выполнению самостоятельной работы

Рекомендуется использовать информационные ресурсы современных электронных библиотек, обращаться к публикациям по в рейтинговых изданиях, сборниках материалов конференций, к записям вебинаров.

Самостоятельная работа включает проработку лекций, чтение обязательной и дополнительной литературы.

При проработке лекций рекомендуется обратиться к конспектам лекционного материала (кратко, схематично, последовательно зафиксированным основным положениям, выводам, формулировкам, обобщениям), проверить использованные в лекции термины и понятия с помощью словарей, справочников с выписыванием толкований в тетрадь.

Следует обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе.

Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии.

При чтении обязательной и дополнительной литературы рекомендуется смысловое чтение, аннотирование и конспектирование. При аннотировании рекомендуется отметить те идеи, положения, которые могут быть использованы в

собственном диссертационном исследовании. При конспектировании – определить цель, ознакомиться с полным текстом источника, определить его логическую структуру, зафиксировать основное содержание структурных компонентов; также рекомендуется фиксировать собственные вопросы, суждения, умозаключения по содержанию конспектируемого источника.

Кроме обязательной и дополнительной литературы, определенной авторами рабочей программы, рекомендуется самостоятельный поиск и проработка дополнительных источников, в том числе аналитического характера (научные статьи, диссертационные исследования).

Также рекомендуется составить перечень основных понятий и терминов (гlossарий) и проанализировать их с использованием словарей, в результате чего должны быть отобраны определения, в наибольшей степени отражающие признаки рассматриваемых явлений.

При подготовке к дифференцированному зачету рекомендуется актуализация и анализ содержания материала лекционных и практических занятий; чтение обязательной и дополнительной литературы; самостоятельный поиск информации по отдельным вопросам с использованием различных видов источников.