

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Географический факультет

УТВЕРЖДАЮ
Декан географического факультета,
академик РАН Добролюбов С.А.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
МЕТОДЫ АНАЛИЗА ГЕОГРАФИЧЕСКИХ ДАННЫХ

Уровень высшего образования:
бакалавриат

Направление подготовки:
05.03.02 «География»

Направленность (профиль) ОПОП:
«Физическая география и ландшафтоведение»

Форма обучения:
очная

Рабочая программа рассмотрена и одобрена
Учебно-методической комиссией географического факультета
(протокол №19, дата 24.03.2023)

Москва 2023

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки «География» (*программы бакалавриата, магистратуры, реализуемые последовательно по схеме интегрированной подготовки*).

ОС МГУ утверждены решением Ученого совета МГУ имени М.В. Ломоносова (приказ по МГУ № 1383 от 30 декабря 2020 года).

Год (годы) приема на обучение: 2021

1. Место дисциплины в структуре ОПОП — относится к вариативной части ОПОП, является обязательной для освоения.
2. Входные требования для освоения дисциплины, предварительные условия: базируется на знаниях по дисциплинам «Математика», «Прикладная математика», «Основы математической статистики и численного моделирования», «Методы физико-географических исследований».
3. Планируемые результаты обучения по дисциплине, соотнесенные с требуемыми компетенциями выпускников

Компетенции выпускников (коды)	Индикаторы (показатели) достижения компетенций	Планируемые результаты обучения по дисциплине, сопряженные с компетенциями
СПК-2.Б (<i>формируется частично</i>) Способен применять полевые, инструментальные, геохимические, геофизические, геоинформационные, аэрокосмические, статистические методы исследования для решения профессиональных задач.	СПК-2.1. Применяет инструментальные, геофизические, геоинформационные, аэрокосмические, статистические методы исследования для решения профессиональных задач.	<p>Знать:</p> <ul style="list-style-type: none"> - сферу применения статистических методов в ландшафтоведении - ограничения каждого статистического метода; <p>Уметь:</p> <ul style="list-style-type: none"> - различать типы географических данных - выбирать адекватный задаче статистический метод анализа данных - разрабатывать алгоритм исследования данных - интерпретировать результаты статистического анализа; <p>Владеть: методами корреляционного, дисперсионного, регрессионного, канонического, кластерного, дискриминантного, факторного анализа.</p>

4. Объем дисциплины 2 зачётные единицы (72 часа), в том числе 36 академических часа на контактную работу обучающихся и 36 академических часов на самостоятельную работу обучающихся.
5. Формат обучения не предполагает электронного обучения и использования дистанционных образовательных технологий (за исключением форс-мажорных обстоятельств – пандемии и т.п.).

6. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических или астрономических часов и виды учебных занятий

Наименование и краткое содержание разделов и тем дисциплины, Форма промежуточной аттестации по дисциплине	Всего (часы)	В том числе							
		Контактная работа (работа во взаимодействии с преподавателем) <i>Виды контактной работы, часы*</i>					Самостоятельная работа обучающегося <i>Виды самостоятельной работы, часы</i>		
		Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Всего	Выполнение практических работ	Всего	
Тема 1. Сфера применения статистических методов анализа данных в ландшафтоведении	2	2				2			
Тема 2. Подготовка географических данных к анализу	6	2	2			4	2	2	
Тема 3. Первичный графический анализ данных и базовые статистики	5	2	1			3	2	2	
Текущая аттестация № 1: контрольная работа	1		1			1			
Тема 4. Задача анализа межкомпонентных связей в ландшафте	18	6	6			12	6	6	
Тема 5. Задача редукции данных	10	2	4			6	4	4	
Тема 6. Задача классификации данных	11	4	3			7	4	4	
Текущая аттестация № 2: контрольная работа	1		1			1			
Промежуточная аттестация – экзамен	18	<i>Устный экзамен</i>							18
Итого	72	36							36

Содержание лекций, семинаров

Содержание лекций

Тема 1. Сфера применения статистических методов анализа данных в ландшафтоведении.

Применимость методов математической статистики к решению задач ландшафтоведения и ландшафтной экологии. Краткая история применения математических методов в ландшафтоведении: возможности, достижения, критика. Детерминированные и вероятностные отношения в ландшафте. Связи-отношения и связи-взаимодействия свойств ландшафта. Статистические гипотезы и ландшафтно-географические гипотезы. Формулировка задачи исследования на ландшафтно-географическом и языке и языке математической статистики. Возможность оценки достоверности результатов. Требования к массовости и репрезентативности данных. Основные типы поисковых задач: анализ связей, классификация, снижение размерности. Принципиальные преимущества и ограничения применимости статистических методов для решения ландшафтно-географических задач. Структура базовых пакетов статистических программ.

Тема 2. Подготовка географических данных к анализу.

Типы географических данных и типы случайных величин. Номинальные, порядковые и количественные переменные. Правила кодирования данных. Правила организации таблиц данных. Особенности географических данных с точки зрения статистической обработки. Требования к типу распределения данных и проверка на нормальное распределение. Способы приведения данных к нормальному распределению. Непрерывные и дискретные распределения. Критерии проверки гипотез. Критерий согласия Пирсона.

Тема 3. Первичный графический анализ данных и базовые статистики.

Способы визуализации данных. Одномерные, двухмерные и трехмерные отображения связей между переменными. Визуальное распознавание типов связей между переменными. Ландшафтный смысл линейных, экспоненциальных, параболических и других нелинейных видов связи. Смысл выбросов. Анализ диапазонов значений функции в разных интервалах значений аргумента. Области запрещенных значений. Смысл информации: ограничение разнообразия состояний. Анализ базовых статистик: проверка на однородность выборки, случайное варьирование или наличие ведущего процесса, принадлежность двух выборок к одной генеральной совокупности. Алгоритм выбора методов в зависимости от типа и распределения данных.

Тема 4. Задача анализа межкомпонентных связей в ландшафте.

Типы задач при анализе связей: между двумя количественными переменными, между двумя качественными переменными, между качественной и количественной переменными, между двумя группами количественных переменных. Алгоритм перебора гипотез о влиянии фактора на исследуемое свойство ландшафта и сужения поля поиска объяснений. Информационный анализ (кросстабуляция) как способ анализа связи между качественными переменными. Параметрические и непараметрические корреляции и условия их применения. Применение дисперсионного анализа и его непараметрических аналогов для исследования отклика свойств ландшафта на категоризованные состояния фактора. Общая оценка значимости категоризованного фактора и анализ попарных различий между категориями состояний фактора. Построение регрессионных уравнений как способ анализа совместного влияния нескольких факторов на исследуемое свойство. Оценка качества уравнения. Выбор значимых предикторов. Интерпретация знака и меры тесноты связи. Сравнение вкладов переменных-предикторов в объяснение варьирования зависимой переменной. Неаддитивные взаимодействия. Оценка вклада нелинейной составляющей связи. Анализ остатков уравнения регрессии.

Тема 5. Задача редукции данных.

Факторный анализ как способ снижения размерности данных и получения новых переменных как линейных комбинаций исходных переменных. Метод главных компонент. Вращение (варимакс-трансформация). Дисперсия и собственные значения. Факторные нагрузки и коэффициенты. Интерпретация новых компонент (факторов). Многомерное шкалирование как непараметрический аналог факторного анализа. Канонические корреляции как способ анализа связи между группами количественных переменных на основе редукции данных.

Тема 6. Задача классификации данных.

Кластерный анализ как способ классификации данных. Деревья классификации. Классификация методом k-средних. Матрица дистанций. Метрики дистанций. Дискриминантный анализ как классификация с обучением. Постериорные вероятности и неопределенность классификации.

План проведения семинаров

1. Обсуждение теоретических основ и ограничений применимости рассматриваемых статистических методов
2. Обсуждение и постановка задачи практической работы
3. Анализ исходных данных и параметров
4. Обсуждение алгоритма решения задачи
5. Выбор статистического метода и реализация
6. Интерпретация результатов

7. Фонд оценочных средств для оценивания результатов обучения по дисциплине:

Текущая аттестация № 1: контрольная работа

Текущая аттестация № 2: контрольная работа

Примерные вопросы для контрольных работ

1. В чем отличия отношений, исследуемых методами математической статистики от алгебраических методов.
2. Какие типовые задачи ландшафтоведения удобно решать статистическими методами, и почему статистические методы адекватны особенностям объектов ландшафтоведения?
3. Что такое номинальные переменные, и какими методами их нельзя обрабатывать?
4. Почему первый этап анализа данных должен заключаться в проверке на соответствие данных нормальному распределению?
5. Что показывает критерий хи-квадрат при проверке гипотезы о нормальности распределения?
6. Какие три описательных статистики должны быть равны или близки при нормальном распределении?
7. Объясните суть асимметрии и эксцесса и нарисуйте соответствующие гистограммы.
8. В каком случае используется t-распределение Стьюдента?

9. Какая задача решается с помощью критерия Фишера?
10. К каким типам переменных применим дисперсионный анализ?
11. Сформулируйте задачу, решаемую дисперсионным анализом на языке географии и на языке статистики.
12. Какие ограничения есть у корреляционного анализа?
13. Что показывает коэффициент детерминации?
14. Для чего используются B и $Beta$ коэффициенты в регрессионном анализе?
15. Зачем применяется вращение в методе главных компонент?
16. От чего зависит выбор метрики дистанции в кластерном анализе?

Примерный перечень практических работ по дисциплине

1. Исследование таблицы данных и выделение типов переменных.
2. Проверка количественных данных на нормальность распределения.
3. Подбор адекватных способов графического изображения для решения поставленной преподавателем задачи.
4. Выдвижение гипотез о причинах межкомпонентных связей по графическим изображениям.
5. Расчет базовых статистик по таблице данных.
6. Обоснование выбора допустимых методов статистического анализа по результатам первичного исследования данных.
7. Корреляционный анализ связи переменных.
8. Проверка значимости геоморфологических классов для объяснения варьирования свойств почв и фитоценозов методом дисперсионного анализа.
9. Построение и сравнение качества линейных и нелинейных регрессионных уравнений.
10. Составление таблицы значимости связей между ярусами фитоценоза, свойствами почв и рельефа по результатам расчета канонических корреляций.
11. Выявление ассоциаций видов растений и/или содержания химических элементов в почвах методами факторного анализа и многомерного шкалирования.
12. Классификация точек полевых наблюдений по схожести свойств растительности, почв и рельефа.

Примерный перечень вопросов для экзамена

1. Основные задачи, решаемые статистическими методами. Виды данных и применяемые к ним методы анализа
2. Детерминированные и вероятностные отношения переменных. Смысл меры ошибки. Смысл уровня значимости
3. Описательные статистики: среднее, мода, медиана, квартиль, стандартное отклонение, дисперсия, асимметрия, эксцесс
4. Нормальное распределение и его свойства. Стандартизация данных
5. Распределения, связанные с нормальным и используемые для принятия решений (хи-квадрат, Стьюдента, Фишера)
6. Основы проверки статистических гипотез. Ошибки первого и второго рода

7. Визуальные методы оценки принадлежности распределения к нормальному
8. Отличия параметрических и непараметрических критериев проверки гипотез
9. Проверка гипотезы нормальности выборки
10. Т-тест для одной и двух выборок
11. Непараметрические аналоги т-критерия Стьюдента для зависимых и независимых выборок
12. Одномерный дисперсионный анализ. Анализ попарных различий
13. Непараметрический однофакторный анализ Краскала-Уоллеса, медианный тест
14. Коэффициент корреляции Пирсона и его непараметрические аналоги
15. Многомерный линейный регрессионный анализ
16. Анализ качества результатов регрессионного анализа: коэффициент детерминации, критерий Фишера, уровень значимости, регрессионные коэффициенты, стандартизованные регрессионные коэффициенты
17. Задачи канонического анализа
18. Факторный анализ и метод главных компонент. Ограничения и интерпретация результатов
19. Кластерный анализ, метод к-средних. Проблема выбора метрики дистанций
20. Задачи дискриминантного анализа

Шкала и критерии оценивания

Промежуточная аттестация по итогам освоения дисциплины – экзамен (в устной форме).

Оценка РО и соответствующие виды оценочных средств	Неудовлетворительно	Удовлетворительно	Хорошо	Отлично
Знания (виды оценочных средств: устный опрос, тесты)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения (виды оценочных средств: практические контрольные задания)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности принципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт)	Отсутствие навыков	Наличие отдельных	В целом, сформированные	Сформированные навыки

<p>деятельности)(виды оценочных средств: практические контрольные задания)</p>		<p>навыков</p>	<p>навыки (владения), но используемые не в активной форме</p>	<p>(владения), применяемые при решении задач</p>
--	--	----------------	---	--

8. Ресурсное обеспечение:

Основная литература:

1. Пузаченко Ю.Г. Математические методы в экологических и географических исследованиях. М.: Академия, 2004. 416 с.

Дополнительная литература:

1. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. 2-е изд. СПб.: Питер, 2003. – 688 стр.
2. Джонгман Р.Г.Г., Тер Браак С.Дж.Ф., Ван Тонгерен О.Ф.Р. Анализ данных в экологии сообществ и ландшафтов. Пер. с англ. под ред. А. Н. Гельфана, Н. М. Новиковой, М. Б. Шадринной. М.: РАСХН, 1999. – 306 с.
3. Дэвис Дж. С. Статистический анализ данных в геологии. М.: Недра, 1990. Т.1. 319 с. Т.2. – 427 с.
4. Халафян А. А. Учебник STATISTICA 6. Статистический анализ данных. М.: Бином-пресс, 2007. – 512 с.

- Перечень лицензионного программного обеспечения
Statistica (от версии 7 и позже)
- Перечень профессиональных баз данных и информационных справочных систем
- реферативная база данных издательства Elsevier: www.sciencedirect.com
- Перечень ресурсов информационно-телекоммуникационной сети «Интернет»
www.statsoft.ru Электронный учебник статистики
- Описание материально-технической базы
Учебная аудитория с мультимедийным проектором
Персональные компьютеры

9. Язык преподавания: русский

10. Преподаватель (преподаватели): Ответственный за курс — Мерекалова Ксения Алексеевна, м.н.с.; преподаватели: Мерекалова Ксения Алексеевна, м.н.с.

11. Разработчики программы: Мерекалова Ксения Алексеевна, м.н.с., Хорошев Александр Владимирович, профессор, д.г.н.