

**Московский государственный университет имени М.В.Ломоносова
Географический факультет**

«Утверждено»
Чл.-корр. РАН С.А. Добролюбов

« _____ » _____ 2018 г.

Согласовано
Учебно-методической комиссией
факультета

« _____ » _____ 2018 г.

ПРОГРАММА ДИСЦИПЛИНЫ

«Методы анализа географических данных»

по направлению подготовки 05.03.02 География,
направленность (профиль) «Физическая география и ландшафтоведение»
уровня высшего профессионального образования бакалавриат
с присвоением квалификации «бакалавр»

1.. Цели и задачи дисциплины

Цель курса - обучение методам статистического анализа ландшафтно-географических данных и интерпретации их результатов.

Задачи:

- Освоение способов постановки исследовательских задач и подготовки данных к анализу
- Освоение статистических методов анализа данных
- Получение навыка интерпретации результатов анализа данных

2. Место дисциплины в структуре ООП:

Дисциплина относится к блоку профильных профессиональных дисциплин вариативной части основной образовательной программы высшего образования по направлению «География», профиль "Физическая география и ландшафтоведение", обязательный курс.

Дисциплина изучается на 3 курсе в 6 семестре.

Дисциплина образует один из важных элементов подготовки студентов в области физической географии. Перечень дисциплин, которые должны быть освоены для начала освоения данной дисциплины: «Ландшафтоведение», «Актуальное ландшафтоведение», «Методы физико-географических исследований», «Прикладная математика».

Освоение данной дисциплины необходимо как предшествующее для приобретения знаний в следующих дисциплинах: «Функционирование, динамика и эволюция ландшафтов», «Физико-математические основы ландшафтоведения», а также для прохождения производственной практики.

3. Требования к результатам освоения дисциплины

Процесс изучения дисциплины направлен на формирование следующих компетенций в соответствии со Стандартом МГУ:

- знание принципов изучения круговоротов вещества, энергии и информации в ландшафте, принципы моделирования процессов и пространственной структуры ландшафта с использованием полевых, геохимических, геофизических, геоинформационных, аэрокосмических методов, методов пространственного анализа (СПК-2; компетенция формируется частично).

В результате освоения дисциплины обучающийся должен:

Знать:

- сферу применения статистических методов в ландшафтоведении
- ограничения каждого статистического метода

Уметь:

- различать типы географических данных
- выбирать адекватный задаче статистический метод анализа данных
- разрабатывать алгоритм исследования данных
- интерпретировать результаты статистического анализа

Владеть:

- методами корреляционного, дисперсионного, регрессионного, канонического, кластерного, дискриминантного, факторного анализа

4. Структура и содержание дисциплины

Объем дисциплины и виды учебной работы.

Общая трудоемкость дисциплины составляет 2 зачетные единицы.

Общая аудиторная нагрузка – 26 часов, в т.ч. семинары – 26 часов.

Объем самостоятельной работы студентов – 46 академических часов.

№	Наименование разделов и тем дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая СРС и трудоемкость (в часах)			Формы текущего контроля успеваемости Форма промежуточной аттестации
				Лекции	Практические занятия (семинары)	СРС	
1	Сфера применения статистических методов анализа данных в ландшафтоведении	6	1	-	2	2	Контрольная работа
2	Подготовка географических данных к анализу	6	2-3	-	4	6	Контрольная работа
3	Первичный графический анализ данных и базовые статистики	6	4-7	-	8	8	Контрольная работа
4	Задача анализа межкомпонентных связей в ландшафте	6	8-13	-	12	14	Контрольная работа. Зачет по практическим заданиям
	ИТОГО			0	26	46	Зачет

5. Содержание дисциплины

План проведения семинаров

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. Представление о полиструктурности и полимасштабности 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. 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.

Тема 3. Первичный графический анализ данных и базовые статистики.

Способы визуализации данных. Одномерные, двумерные 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 непрерывными переменными, между двумя дискретными переменными, между дискретной и непрерывной переменными, между двумя группами непрерывных переменных. Алгоритм перебора гипотез о влиянии фактора на исследуемое свойство ландшафта и сужения поля поиска объяснений. 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. Построение регрессионных уравнений межкомпонентных связей в скользящем квадрате. Разграничение ареалов с единым типом межкомпонентных отношений. Канонические корреляции как способ анализа связи между группами количественных переменных. Информационный анализ как способ анализа связи между качественными переменными.

Перечень практических заданий

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. Последовательность применения методов для анализа межкомпонентных связей по теме курсовой работы.
8. Корреляционный анализ связи переменных.
9. Проверка значимости геоморфологических классов для объяснения варьирования свойств почв и фитоценозов методом дисперсионного анализа.
10. Построение и сравнение качества линейных и нелинейных регрессионных уравнений.

11. Составление таблицы значимости связей между ярусами фитоценоза, свойствами почв и рельефа по результатам расчета канонических корреляций.

6. Учебно-методическое обеспечение самостоятельной работы студентов по дисциплине

Задания для самостоятельной работы.

1. Оценить применимость статистических методов к решению задач курсовой работы.
2. Построить и предложить интерпретацию кластер-диаграмм.
3. Рассчитать вероятности соответствия сочетаний свойств почв и фитоценозов типичным значениям для классов рельефа, рассчитать меру неопределенности.
4. Рассчитать значения виртуальных факторов и дать интерпретацию их экологического смысла.

7. Оценочные средства для текущего контроля успеваемости

Примерные вопросы для контрольных работ

1. Основные понятия статистики
2. Случайные величины и функции распределения. Плотность вероятности
3. Выборочные характеристики Описательные статистики
4. Дискретные и непрерывные распределения.
5. Нормальное распределение и его свойства
6. Логнормальное распределение
7. Гамма распределение
8. Стратегии работы с данными в зависимости от принадлежности выборки к тому или иному распределению
9. Распределения, связанные с нормальным и используемые для принятия решений (хи-квадрат, Стьюдента, Фишера)
10. Основы проверки статистических гипотез.
11. Ошибки первого и второго рода.
12. Понятие генеральной совокупности и выборки
13. Параметры распределений. Параметры выборки. Оценивание параметров распределения по выборке
14. Проверка гипотезы нормальности выборки.
15. Визуальные методы оценки принадлежности распределения к нормальному
16. Проверка гипотезы о тождественности параметров двух независимых выборок при допущении, что обе выборки нормальны
17. Непараметрические критерии проверки гипотез
18. Одномерный дисперсионный анализ
19. Непараметрический однофакторный анализ Краскла, медианный тест
20. Представление о многомерном пространстве
21. Коэффициент корреляции Пирсона. Ошибка коэффициента корреляции.
22. Непараметрические коэффициенты корреляции
23. Модель линейного регрессионного анализа. Простая линейная регрессия
24. Метод наименьших квадратов
25. Связь регрессии с корреляцией.
26. Многомерный регрессионный анализ

27. Методы прямого и обратного пошагового выбора
28. Анализ качества результатов регрессионного анализа

Зачет по практической работе выставляется по результатам выполнения работы после исправления указанных преподавателем недостатков и ответа на замечания и вопрос»

8. Формы и содержание промежуточной аттестации

Примерный перечень вопросов к устному экзамену

1. Основные задачи, решаемые статистическими методами.
2. Описательные статистики: среднее, мода, медиана, квартиль, стандартное отклонение, дисперсия, асимметрия, эксцесс
3. Нормальное распределение и его свойства
4. Стратегии работы с данными в зависимости от принадлежности выборки к тому или иному распределению
5. Основы проверки статистических гипотез.
6. Визуальные методы оценки принадлежности распределения к нормальному
7. Проверка гипотезы о тождественности параметров двух независимых выборок при допущении, что обе выборки нормальны
8. Непараметрические критерии проверки гипотез
9. Одномерный дисперсионный анализ
10. Непараметрический однофакторный анализ Краскла, медианный тест
11. Коэффициент корреляции Пирсона и его непараметрические аналоги
12. Многомерный регрессионный анализ
13. Методы прямого и обратного пошагового выбора
14. Анализ качества результатов регрессионного анализа
15. Задачи кластерного анализа
16. Задачи дискриминантного анализа
17. Задачи факторного анализа

9. Учебно-методическое и информационное обеспечение дисциплины

а) Основная литература

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

б) Дополнительная литература

Дэвис Дж. С. Статистический анализ данных в геологии. М.: Недра, 1990. Т.1. 319 с. Т.2. 427 с.

Дмитриев Е.А. Математическая статистика в почвоведении. М.: ЛИБРОКОМ, 2009. 328 с.

Боровиков В. П., Боровиков И. П. STATISTICA Статистический анализ и обработка данных в среде Windows. М.: Информационно-издательский дом ФИЛИНЪ, 1997, 592 с.

10. Материально-техническое обеспечение дисциплины

А. Учебная аудитория с персональными компьютерами на 15 мест

Б. Мультимедиа-проектор.

11. Контролирующие материалы по дисциплине (ФОС)

Тесты контроля остаточных знаний по дисциплине

1. Какие свойства данных определяют выбор параметрических или непараметрических методов анализа? а) количество данных не менее 100, б) наличие линейных отношений между переменными, в) **соответствие нормальному распределению**
2. Что входит в число ограничений применения корреляционного анализа? а) **линейный характер связи**, б) отклонение распределения данных от логнормального, в) **отклонение распределения данных от нормального**, г) **количественный характер переменных**
3. Какой статистический метод позволяет отразить вид связи между переменными: а) **регрессионный анализ**, б) корреляционный анализ, 3) дисперсионный анализ, 4) дискриминантный анализ
4. Какова область значений коэффициента корреляции? а) **от -1 до 1**, б) от 0 до 1, в) от 0 до 100, г) от минус бесконечности до плюс бесконечности

Программа составлена в соответствии с требованиями Образовательного стандарта МГУ по направлению подготовки 05.03.02 «География», профиль «Физическая география и ландшафтоведение»

Программа одобрена на заседании кафедры физической географии и ландшафтоведения

Протокол № 8 от 11 декабря 2018 г.

Заведующий кафедрой _____ Дьяконов К.Н.

Разработчики:

Хорошев Александр Владимирович	доцент, д.г.н.	МГУ имени М.В.Ломоносова,
Меркалова Ксения Алексеевна	м.н.с.	географический факультет, кафедра физической географии и ландшафтоведения

Эксперт:

В.В.Сысуев, д.г.н., профессор,	МГУ имени М.В.Ломоносова, географический факультет, кафедра физической географии и ландшафтоведения
---------------------------------------	--