

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

УТВЕРЖДАЮ

и.о. декана П.В.Красильников /_____ /

«__» _____ 20__ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:

АНАЛИЗ ДАННЫХ

Уровень высшего образования:

Магистратура

Направление подготовки (специальность):

06.04.02 Почвоведение

Направленность (профиль) ОПОП:

Агрохимия и агроэкология

Форма обучения: очная

Рабочая программа рассмотрена и одобрена учебно-методической комиссией
факультета почвоведения (протокол № _____, дата _____)

Председатель УМК _____ Рахлеева А.А.

Москва 2021

Рабочая программа дисциплины «Анализ данных» разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 06.04.02. Почвоведение программы магистратуры

ОС МГУ утвержден решением Ученого совета МГУ имени М.В.Ломоносова от 28 декабря 2020 года (протокол № 7).

1. Место дисциплины в структуре ОПОП: Дисциплина входит в вариативную часть ОПОП и является обязательной для освоения.

2. Входные требования для освоения дисциплины, предварительные условия: для освоения дисциплины необходимы базовые знания, полученные в бакалавриате по дисциплинам: иностранный язык (английский), ботаника, учение о почвенных свойствах и процессах, русский язык и культура речи, безопасность жизнедеятельности, общая химия, аналитическая химия, органическая химия, математическая статистика, экология, ботаника с основами геоботаники, модуль «Почвоведение», модуль «Науки о почве».

3. Планируемые результаты обучения в результате освоения дисциплины, соотнесенные с требуемыми компетенциями выпускников:

<p>Компетенции выпускников (коды) М-СПК-4.</p> <p>Владеет методами исследования, логического и статистического анализа и отображения данных о пространственном варьировании объектов агроценоза.</p>	<p>Индикаторы: М-СПК-4.1. Использует методы исследования, логического и статистического анализа и отображения данных о пространственном варьировании объектов агроценоза.</p>	<p>Планируемые результаты обучения по дисциплине, сопряженные с компетенциями</p> <p>Знать: - современные методы пилотной обработки почвенных и экологических данных, стандартный набор приемов статистической обработки - влияние особенностей почвенного пространства на результаты</p> <p>Уметь: - выявлять ошибки, -гармонизировать данные, -оценивать правильность экспериментальных планов интерпретировать результаты разных схем дисперсионного анализа почвенных данных</p> <p>Владеть: - методами выявления ошибок в данных, понимать возможности их интерпретации и исправления</p> <p>Иметь опыт деятельности: - в сфере Data Mining. -в сфере графической обработки представленных данных, - в сфере работы со статистическими программами</p>
--	---	--

4. **Объем дисциплины** 2 з.е., в том числе 24 академических часов на контактную работу обучающихся с преподавателем, 48 академических часа на самостоятельную работу обучающихся.

5. **Формат обучения** очный (с применением электронного обучения и дистанционных образовательных технологий) *(отметить, если дисциплина или часть ее реализуется с использованием электронного обучения и дистанционных образовательных технологий)*

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

Наименование и краткое содержание разделов и тем дисциплины / форма текущей аттестации	Всего (часы)	В том числе							
		Контактная работа (работа во взаимодействии с преподавателем)				Самостоятельная работа обучающегося			
		Занятия лекционного типа	Занятия семинарского типа (семинары)	Занятия семинарского типа (лабораторные)	Занятия семинарского типа (практические)	Всего	Написание рефератов	Подготовка докладов	Всего
Раздел 1. Особенности анализа почвенных данных/тестирование.	12	2	2			4	4	4	8
Раздел 2. Планирование эксперимента и дисперсионный анализ/устный опрос, доклад с презентацией	20	4	4			8	4	8	12
Раздел 3. Методы Data Mining/доклад с презентацией	12	4	4			8		14	14
Раздел 4. Обзор характерных ошибок при использовании статистических методов/устный вопрос,	16	2	2			4	12		12

тестирование										
		12	12			24	20	26	46	
Промежуточная аттестация зачет		Зачет								
ИТОГО	72	24					48			

Подробное содержание разделов и тем дисциплины:

Раздел 1. Особенности анализа почвенных данных

Тема 1.1 Типология задач в почвоведении и экологии. Элемент опробования. Способы опробования. Схемы площадного опробования. Достоинства и недостатки регулярных и случайных схем. Комбинированные схемы. Влияние неоднородности природных сред и характеристик элементов опробования на статистические характеристики

Тема 1.2. Понятие «данные». Шкалы измерений в экологии и почвоведении. Случайная величина как модель порождения данных. Многомерные данные. Пространственно распределенные данные. Закономерные и случайные составляющие в пространственной изменчивости природных сред.

Раздел 2. Планирование эксперимента и дисперсионный анализ.

Тема 2.1. Основные свертки количественных данных. Квантильный анализ. Среднее, дисперсия, коэффициент вариации свойств. Эксперименты в науке и промышленности. Различия в методике. Общие идеи планирования эксперимента. Вычислительные проблемы. Анализ мощности и оценка объема выборки при планировании эксперимента. Примеры планирования эксперимента в экологии и почвоведении.

Тема 2.2. Схемы дисперсионного анализа. Дисперсионный анализ (ДА). Однофакторный ДА. Главные эффекты ДА. Факторный ДА. Гнездовые планы ДА. Сбалансированный ДА. Оценка компонент дисперсии. Пример интерпретации результатов дисперсионного анализа данных динамики почвенных свойств.

Раздел 3. Методы Data Mining

Тема 3.1. Разведочный анализ данных как метод нахождения связей между переменными. Основные методы. Возможности графического анализа в пакете STATISTICA. Матричные и спектральные графики. Пиктографики. Анализ нечисловой информации. Таблицы сопряженности. Многовходовые таблицы. Логлинейный анализ в таблицах частот

Тема 3.2. Применение кластерного анализа в почвоведении и экологии. Выдвижение гипотез о структуре данных. Интерпретационные возможности различных модификаций кластерного анализа. Факторный анализ. Решаемые задачи. Вычислительные процедуры. Интерпретация. Пример факторного анализа для анализа эволюции почвенного покрова. Дискриминантный анализ при оценке состояния почвенного покрова. Примеры.

Тема 3.3 Обобщенные линейные модели. Возможности. Условия применимости. Проверка моделей. Логистическая регрессия. Примеры использования этого подхода для анализа данных о распространенности тех или иных видов растительности. Нелинейное оценивание. Основные идеи. Примеры применения.

Раздел 4. Обзор характерных ошибок при использовании статистических методов.

Тема 4.1. Типовые ошибки в применении статистических методов анализа данных. Ошибки в обозначениях. Ошибки в расчетах параметров. Ошибки представления расчетов. Ошибки в формулировке нулевых гипотез. Интерпретация нулевых и

альтернативных гипотез. Ошибки статистического вывода. Мощность критерия как индикатор интерпретационных возможностей статистического вывода.

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

7.1. Типовые контрольные задания или иные материалы для проведения текущего контроля:

Рекомендуемые темы докладов:

1. Приемы выделения закономерной составляющей в пространственной изменчивости почвенных свойств
2. Способы планирования эксперимента в почвоведении
3. Что такое Data Mining?
4. Примеры использования нелинейной регрессии в почвоведении и агрохимии
5. Анализ статистических приемов в статье по теме квалификационной работы
6. Приемы графического анализа первичной информации
7. Обзор русскоязычных статистических программ
8. Корреляционный анализ как инструмент выявления закономерностей

7.2. Типовые контрольные вопросы, задания или иные материалы для проведения промежуточной аттестации:

1. Что такое элемент опробования?
2. Насколько сильно может влиять схема опробования на конечные результаты?
3. Какие схемы опробования Вы знаете?
4. Роль исследователя в формировании представления о почве и почвенном покрове
5. Какие шкалы измерений используются в почвоведении?
6. Каким образом можно разделить закономерные и случайные составляющие в пространственной изменчивости природных сред?
7. Какие основные свертки экспериментальных данных Вы знаете?
8. Что такое мощность критерия?
9. Как оценить объем выборки для эксперимента?
10. Что такое дисперсионный анализ?
11. Что такое «Факторный дисперсионный анализ»?
12. Что такое «Главные эффекты»?
13. Что такое «Взаимодействие»?
14. Что такое иерархическая схема дисперсионного анализа?
15. Каким образом можно оценить компоненты дисперсии?
16. Что такое «разведочный» (пилотный) анализ данных?
17. Какие графические методы анализа реализованы в пакетах STATISTICA и Excel?
18. Что такое матричные графики?
19. Что такое Пиктографики?
20. Что такое таблицы сопряженности?
21. Какие типы таблиц сопряженности Вы знаете?
22. Что такое логлинейный анализ?
23. Типы кластерного анализа, реализованные в пакете STATISTIC?
24. Как интерпретируются результаты иерархического кластерного анализа?
25. Что такое факторный анализ?
26. Как используется дискриминантный анализ в почвоведении?
27. Что такое нелинейное оценивание?
28. Где применяется логистическая регрессия?
29. Что такое «обобщенные линейные модели»?

30. Условия применимости многомерного регрессионного анализа в почвоведении
31. Типы статистических ошибок в статьях и монографиях?
32. Минимальные необходимые сведения об эксперименте?

8. Шкала и критерии оценивания результатов обучения по дисциплине:

В таблице представлена шкала оценивания результатов обучения по дисциплине. Уровень знаний обучающегося оценивается на "отлично", "хорошо", "удовлетворительно", "неудовлетворительно".

Оценка "отлично" выставляется, если обучающийся демонстрирует сформированные систематические знания, умения и навыки их практического использования. Оценка "хорошо" ставится, если при демонстрации знаний, умений и навыков студент допускает отдельные неточности (пробелы, ошибочные действия) непринципиального характера. При несистематических знаниях, демонстрации отдельных (но принципиально значимых навыков) и затруднениях в демонстрации других навыков выставляется оценка «удовлетворительно». Оценка "неудовлетворительно" ставится, если знания и умения фрагментарны, а навыки отсутствуют.

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка РО и соответствующи е виды оценочных средств	2	3	4	5
Знания <i>устные и письменные опросы</i>	Отсутстви е знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированны е систематические знания
Умения <i>практические контрольные задания</i>	Отсутстви е умений	В целом успешное, но не систематическо е умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиальног о характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) <i>Презентация по анализу выбранной статьи</i>	Отсутстви е навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированны е навыки (владения), применяемые при решении задач

9. Ресурсное обеспечение:

Перечень основной литературы

1. Дмитриев Е.А. Теоретические и методологические проблемы почвоведения. М., 2001, ГЕОС. 374 стр. шифр ЗУБ, Д-533
2. Самонова В.П. Пространственная изменчивость свойств почв (на примере дерново-подзолистой почвы. Москва: URSS Изд-во ЛКИ, 2008. 156 с.

3. Дмитриев Е.А. Математическая статистика в почвоведении. Москва: URSS : ЛИБРОКОМ, 2009. 326стр. Шифр ЗУБ, Д-533
4. Доспехов Б.А. Борис Александрович . Методика полевого опыта : (с основами стат. обраб. результатов исслед.) : учеб. для студентов высш. с.-х. учеб. заведений М. : Альянс, 2011. 351 с. Шифр: 6ВБ, Д-705
5. Пифо Х.П. Статистика. Москва, 2011, Изд. ВНИИА. 296 стр. Шифр: 6, П-357
6. Боровиков В. STATISTICA. Искусство анализа данных на компьютере. С-Пб, Питер, 2001. 656 стр. Шифр: 3Ф, А-640
7. Статья по выбору из журнала Почвоведение, Агрохимия и т.п.

Перечень лицензионного программного обеспечения: Microsoft office, STATISTICA,
Перечень профессиональных баз данных и информационных справочных систем:
<https://nbmgu.ru/>

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:
<http://statsoft.ru>,

Описание материально-технической базы: компьютерный класс

10. **Язык преподавания:** русский

11. **Преподаватель**

Самсонова Вера Петровна

Должность доцент

Ученая степень доктор биологических наук (по специальности 03.00.27 почвоведение, Диссовет Д501.001.57, МГУ имени М.В. Ломоносова 18 апреля 2003 г. (утв. ВАК)),

Ученое звание доцента по специальности почвоведение (биологические науки) - 26 ноября 1997 г. Приказ Высшей Аттестационной Комиссии

12. **Разработчики программы:**

Самсонова Вера Петровна

Должность доцент

Ученая степень доктор биологических наук (специальности 03.00.24 – почвоведение, Диссовет Д501.001.57, МГУ имени М.В. Ломоносова 18 апреля 2003 г. (утв. ВАК)),

Ученое звание доцента по специальности *почвоведение* (биологические науки) - 26 ноября 1997 г. Приказ Высшей Аттестационной Комиссии

13. **Краткая аннотация дисциплины:**

В курсе АНАЛИЗ ДАННЫХ студенты получают развернутые знания по методам пилотной и более продвинутой обработки данных их области почвоведения и агрохимии. Акцентируется внимание на роли исследователя в процессе получения первичной информации, приводятся примеры влияния способа опробования на конечные результаты, разбираются приемы планирования почвенного эксперимента, обсуждается мощность статистических критериев и подходы к выбору адекватных методов статистического анализа почвенной и экологической информации. Большое внимание уделяется графическим методам анализа и типовым ошибкам в применении статистических методов. Рассматриваются методы сравнительно новой области статистического анализа Data Mining.