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:
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Уровень высшего образования:
Бакалавриат

Направление подготовки (специальность):
06.03.02 Почвоведение

Направленность (профиль) ОПОП:

Форма обучения: очная

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

Москва 2021

На обратной стороне титула:

Рабочая программа дисциплины «Математическая статистика» разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 06.03.02. Почвоведение программы бакалавриата

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 базовую 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: информатика, высшая математика, иностранный язык (английский), учение о почвенных свойствах и процессах, общая химия, аналитическая химия, органическая химия, экология, ботаника и ботаника с основами геоботаники.

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

Компетенции выпускников (коды)	**Индикаторы (показатели) достижения компетенций**	**Планируемые результаты обучения по дисциплине, сопряженные с компетенциями**
Б-ОПК-1	**ОПК-1.1** Применяет знания основных общих закономерностей в области математики, физики, химии, наук о Земле, биологии и экологии для решения профессиональных задач	***Знать:*** - организацию полевого и лабораторного эксперимента; - отличия выборки и генеральной совокупности, а также объекта исследования и элемента опробования; - методы обработки единичной выборки; - однофакторный и двухфакторный дисперсионный анализ; - корреляционный и регрессионный анализы; - кластерный анализ; - метод главных компонент; - дискриминантный анализ. ***Уметь:*** - строить гистограмму, коробочку с усиками, график квантилей, график на нормальной вероятностной бумаге; - выбрать подходящий вид анализа для своих данных; - рассчитать объем выборки в зависимости от предполагаемой нулевой гипотезы; - проводить отбраковку данных, у которых предполагается нормальное распределение; - работать со статическими программами. ***Владеть:*** - навыками первичного анализа данных; - знаниями, необходимыми для обработки данных своего эксперимента; - основными видами анализа данных. ***Иметь опыт деятельности:*** - работы с модулем Анализ данных в Excel, программе Statistica и программной среде R - опыт работы с разными типами шкал (номинальные, ординальные, интервальные

		и абсолютные).

4. **Объем дисциплины** 4 з.е.**,** в том числе 120 академических часов на контактную работу обучающихся с преподавателем (60 академических часов лекционных, 30 академических часов семинаров, 30 академических часов практических занятий), 24 академических часа на самостоятельную работу обучающихся.

5. **Формат обучения** очный с применением 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 (практические)	**Всего**	Заполнение рабочей тетради и выполнение практических занятий	**Всего**
Раздел 1. Случайная величина как математическая модель	**30**	**12**	**10**	**6**	**28**	**2**	**2**
Тема 1.1. Первичная обработка данных	6	4	2	0	6	0	0
Тема 1.2. Понятие статистического испытания	6	2	2	2	6	0	0
Тема 1.3. Константы и параметры распределений	6	2	2	2	6	0	0
Тема 1.4. Законы распределений	12	4	4	2	10	2	2
Форма текущей аттестации по разделу – проверка рабочей тетради, сдача практических работ и устный опрос							
Раздел 2. Классические методы анализа единичной выборки.	**32**	**12**	**10**	**6**	**28**	**4**	**4**
Тема 2.1. Понятие об испытаниях, событиях и величинах.	6	2	2	2	6	0	0
Тема 2.2. Точечные и интервальные оценки	12	4	4	2	10	2	2
Тема 2.3. Статистические гипотезы и их проверка	14	6	4	2	12	2	2

Форма текущей аттестации по разделу – проверка рабочей тетради, сдача практических работ и устный опрос							
Раздел 3. Анализ группы выборок (дисперсионный, корреляционный, регрессионный)	**28**	**10**	**8**	**6**	**24**	**4**	**4**
Тема 3.1. Дисперсионный анализ	12	4	4	2	10	2	2
Тема 3.2 Корреляционный и регрессионный анализы	16	6	4	4	14	2	2
Форма текущей аттестации по разделу – проверка рабочей тетради, сдача практических работ и устный опрос							
Раздел 4. Непараметрические методы анализа данных	**4**	**2**	**2**	**0**	**4**	**0**	**0**
Тема 4.1. Непараметрические методы	4	2	2	0	4	0	0
Форма текущей аттестации по разделу – проверка рабочей тетради, сдача практических работ и устный опрос							
Промежуточная аттестация	2	***экзамен***				**2**	**2**
Раздел 5. Многомерные случайные величины	**10**	**6**		**2**	**8**	**2**	**2**
Тема 5.1. Понятие об испытании в многомерном статистическом анализе	6	4		2	6	0	0
Тема 5.2. Теоретические основы многомерных наблюдений	4	2		0	2	2	2
Форма текущей аттестации по разделу – проверка рабочей тетради, сдача практических работ и устный опрос							
Раздел 6. Исследование зависимостей в случае многомерных данных	**18**	**10**		**4**	**14**	**4**	**4**
Тема 6.1. Обобщение одномерных дисперсионных моделей на многомерный случай	8	4		2	6	2	2
Тема 6.2. Множественный корреляционный анализ	4	2		0	2	2	2
Тема 6.3. Обобщение одномерных регрессионных моделей на многомерный случай	6	4		2	6		0
Форма текущей аттестации по разделу – проверка рабочей тетради, сдача практических работ и							

устный опрос							
Раздел 7. Методы численной классификации	**18**	**8**		**6**	**14**	**4**	**4**
Тема 7.1. Сходство и различие почвенных объектов	4	2		0	2	2	2
Тема 7.2. Кластерный анализ.	6	2		2	4	2	2
Тема 7.3. Метод главных компонент	4	2		2	4	0	0
Тема 7.4. Понятие о дискриминантном анализе	4	2		2	4	0	0
Форма текущей аттестации по разделу – проверка рабочей тетради, сдача практических работ и устный опрос							
Промежуточная аттестация	2	***зачет***				**2**	**2**
ИТОГО	**144**	**120**				**24**	

Подробное содержание разделов и тем дисциплины:

Раздел 1. Случайная величина как математическая модель

Тема 1.1 Первичная обработка данных

Определение понятия «данные». Основные шкалы измерений. Иерархия шкал.

Обработка данных на разных шкалах. Ряды распределений. Классы и Группировка. Абсолютные и относительные частоты. Графическое представление распределений. Гистограммы качественных, порядковых и количественных признаков. Квантильное представление данных. Интерпретация результатов.

Тема 1.2. Понятие статистического испытания

Вероятность. Устойчивость частот. Вероятность. События невозможные и достоверные. Несовместимые события. Независимость событий. Распределение вероятностей для дискретных величин. Распределение непрерывных случайных величин. Плотность вероятности. Интеграл вероятности.

Тема 1.3. Константы и параметры распределений

Мода. Медиана. Квантили. Среднее арифметическое. Центральные отклонения и свойства среднего. Среднее функций от случайных величин. Правильность и систематические ошибки почвенно-агрохимических анализов. Дисперсия. Стандартное отклонение и его свойства. Стандартное отклонение функций от случайных величин. Нормированное отклонение и его особенности. Моменты.

Тема 1.4. Законы распределений

Законы распределений дискретных величин (биномиальный и закон Пуассона). Нормальный и логарифмически нормальный законы распределений.

Раздел 2. Классические методы анализа единичной выборки

Тема 2.1. Понятие об испытаниях, событиях и величинах

Величины случайные и детерминированные. Общие и второстепенные условия проведения испытаний. Объект исследования. Физическая совокупность и ее компоненты. Статистическая совокупность. Генеральные совокупности. Объем совокупностей. Многомерные случайные величины. Случайная величина как математическая модель.

Понятие о выборке. Репрезентативность и рандомизация. Механический отбор. Таблица случайных чисел и ее использование для получения рандомизированной выборки. Послойная выборка. Значение рандомизации.

Тема 2.2. Точечные и интервальные оценки

Точечные оценки констант и параметров. Качество оценок. Несмещенность, эффективность, состоятельность. Робастность. Оценки моды, медианы и среднего арифметического. Характеристика вариабельности случайных величин с помощью лимитов и размахов варьирования. Оценка дисперсии. Число степеней свободы. Оценка стандартного отклонения по размаху. Усреднение оценок дисперсий. Оценка дисперсии функций случайных величин. Оценки других констант распределений. Оценка коэффициентов вариации. Асимметрия и эксцессивность и их характеристика. Ошибки репрезентативности. Ошибка среднего. Ошибки функций от выборочных оценок среднего. Ошибка доли. Интервальные оценки среднего и дисперсии. Точность опыта и погрешность оценки среднего. Гарантированные минимум и максимум среднего. Оценка интервала возможных значений признака. Интервальная оценка доли.

Тема 2.3 Статистические гипотезы и их проверка

Понятие о статистической гипотезе. Алгоритм проверки гипотез. Нулевая и альтернативные гипотезы. Понятие о статистическом критерии. Доверительная вероятность и уровень значимости. Ошибки первого и второго родов.

Статистический анализ единичной выборки. Выбраковка. Аппроксимирующие распределения и критерии согласия. Проверка нормальности распределения. Критерий хи-квадрат. Критерий Колмогорова-Смирнова. Сравнение средних с постоянными величинами. Анализ двух выборок. Сравнение дисперсий. Сравнение средних. Средняя разность и ее значимость. Сравнение средних при одинаковости дисперсий. Сравнение средних при отличающихся дисперсиях. Сравнение долей. Планирование необходимого объема выборок для сравнения средних.

Раздел 3. Анализ группы выборок (дисперсионный, корреляционный, регрессионный)

Тема 3.1 Дисперсионный анализ

Дисперсионный анализ. Постановка задачи. Одно и двухфакторные дисперсионные комплексы. Фиксированные и случайные градации факторов. Первая, вторая и смешанная модели дисперсионных комплексов. Равномерные и неравномерные комплексы. Условия применимости дисперсионного анализа. Разложение дисперсий при дисперсионном анализе. Техника дисперсионного анализа на примере однофакторного анализа. Сравнение группы средних. Наименьшая существенная разность (НСР). Интерпретация результатов дисперсионного анализа.

Тема3.2. Корреляционный и регрессионный анализы

Корреляционный анализ. Связи функциональные и корреляционные. Графическое изображение связей. Степень, форма, направление связей. Коэффициент корреляции и коэффициент детерминации. Оценка значимости коэффициента корреляции.

Регрессионный анализ. Связь дисперсионного и регрессионного анализов. Эмпирическая и теоретическая линии регрессии. Уравнение регрессии. Коэффициенты регрессии и их смысл. Линейная регрессия. Статистическая значимость параметров регрессии и ее оценка. Доверительная зона регрессии. Связь коэффициентов регрессии и коэффициента корреляции.

Раздел 4. Непараметрические методы анализа данных

Тема 4.1. Непараметрические методы

Квантильный анализ одной и более группы выборок. Проверка гипотез об однородности (серийный критерий, критерий Колмогорова-Смирнова, критерий Вилкоксона).

Ранговый коэффициент корреляции Спирмена. Тетрахорический показатель связи. Несимметричные меры ассоциации. Коэффициент конкордации.

Раздел 5. Многомерные случайные величины

Тема 5.1. Понятие об испытании в многомерном статистическом анализе

Отличие одномерных и многомерных случайных величин. Представление данных в формальном виде. Примеры объектов и примеры признаков в почвоведении и экологии. Особенности данных в почвенных и экологических исследованиях. Этапы анализа исследуемой реальной системы. План сбора исходной информации в почвоведении. Представление данных в формальном виде. Первичная обработка данных. График

«коробочка с усиками». Нормальная вероятностная бумага. Что делать с пропущенными значениями. Процесс проведения научного исследования. Схемы пробоотбора. Основные этапы статистической обработки данных. Понятие о мнимых повторностях. Возможности экстраполяции результатов исследования. Понятие об анализе временных рядов. Понятие о пространственном анализе. Основные разделы многомерного анализа данных.

Тема 5.2. Теоретические основы многомерных наблюдений

Случайный вектор. Случайная матрица и ее реализация. Вектор средних как аналог среднего арифметического. Понятие ковариации. Ковариационная матрица, как аналог дисперсии, и ее свойства. Диагональная матрица ковариации. Пример из экологии реализации 3-х мерной случайной матрицы, вектора средних и ковариационной матрицы.

Многомерное нормальное наблюдение и его свойства. Линейная комбинация нормально распределенных случайных векторов, ее среднее и дисперсия. Классификация типов данных и выбор способа анализа зависимостей. Результирующие (зависимые) и объясняющие (независимые) переменные. Основные типы исследования зависимостей. Вероятность превышения p при проверке статистических гипотез.

Раздел 6. Исследование зависимостей в случае многомерных данных

Тема 6.1 Обобщение одномерных дисперсионных моделей на многомерный случай

Роль Р. Фишера. Разложение дисперсий при дисперсионном анализе. Многофакторные дисперсионные комплексы. Нулевая гипотеза многомерного дисперсионного анализа. Двухфакторный дисперсионный анализа без взаимодействия. Условия применимости дисперсионного анализа и использование нелинейных преобразований исходных данных. Критерии для сравнения группы средних: НСР, Ньюмена-Кольса, Шеффе, Тьюки. Пример реализации двухфакторного дисперсионного анализа в программе Statistica 10. Пример применения к одному набору данных разных дисперсионных комплексов: однофакторных, двухфакторного, трехфакторного.

Тема 6.2 Множественный корреляционный анализ

Связи функциональные и корреляционные. Графическое отображение многомерных связей. Обобщение коэффициент корреляции Пирсона на многомерный случай. Связь ковариационной матрицы или корреляционной матрицы. Оценка значимости коэффициента корреляции. Множественный коэффициент корреляции. Коэффициент детерминации, его свойства и оценка. Скорректированный коэффициент детерминации. Частный коэффициент корреляции, его свойства и примеры.

Тема 6.3 Обобщение одномерных регрессионных моделей на многомерный случай

Цели регрессионного анализа. Уравнение множественной линейной регрессии. Нелинейная регрессия. Оценка параметров уравнения регрессии. Гипотеза об адекватности модели. Гипотеза о влиянии переменной на предсказание Y. Статистическая значимость отличия параметров от нуля. Построение доверительных интервалов для коэффициентов регрессии. Гипотеза о равенстве коэффициента регрессии числу. Регрессия от стандартизированных переменных. Анализ остатков. Допустимые преобразования независимых переменных при наличии положительной асимметрии. Сравнение дисперсионных и регрессионных моделей для одних и тех же данных. Понятие о мультиколлинеарности. Критерии для оценки качества подбора регрессионных моделей. Пример реализации регрессионного анализа в программе Statistica 10.

Раздел 7. Методы численной классификации

Тема 7.1 Сходство и различие почвенных объектов

Понятие расстояния. Евклидово расстояние для двухмерного случая. Евклидово расстояние для многомерного случая. Расстояние Манхеттен. Обобщенный вид расстояния для количественных признаков. Масштабный коэффициент. Процент несогласия. Веса почвенных признаков. Расстояние Махаланобиса. Понятие сходства объектов. Среднее внутригрупповое сходство объектов. Среднее межгрупповое сходство объектов. Понятие о методах объединения: метод ближнего соседа, метод дальнего соседе, центроидный метод, метод среднего сходства, метод Варда.

Тема 7.2 Кластерный анализ

Определение кластерного анализа. Иерархические схемы классификации и дивизивные методы. Построение дендрограмм и их интерпретация. Метод к-средних. Примеры реализации в программе Statistica 10. Принципы сравнение почвенных профилей. Примеры использования кластерного анализа в почвоведении: мультисубстратное тестирование почвенных микробных сообществ, кластерный анализ почв парков г. Москвы по формуле профиля.

Тема 7.3 Метод главных компонент

Понятие о главной компоненте. Суть метода. Геометрическая интерпретация. Задачи метода. Пример преобразования корреляционной матрицы. Собственные числа. График каменистой осыпи. Собственные вектора. Проекция переменных на факторную плоскость. Проекция объектов на главные компоненты. Пример применения в почвоведении. Понятие о факторном анализе.

Тема 7.4 Понятие о дискриминантном анализе

Задачи дискриминантного анализа. Примеры распознавания образов для почвенных и экологических данных. Формальная постановка для случая дискретных классов. Дискриминантная функция. Статистические и эвристические методы. Правила классификации. Оценивание ошибки классификации. Процедура классификации и «отказ от классификации». Лямбда Уилкса. Толерантность. Примеры реализации в программе Statistica 10. Пример картографирование почвенного покрова и засоленности с использованием дискриминантного анализа.

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

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

Типовое задание №1

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) Выписать признаки и указать для них тип шкалы; 2) построить ранжированный ряд; 3) для данных рассчитать квантили (нижний и верхний децили, нижний и верхний квантили, медиану), определить минимум и максимум; 4) построить огиву; 5) построить график квантилей.

Типовое задание №2

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) группировка количественных данных; 2) группировка категоризированных данных; 3) построить гистограмму и полигон частот; 4) построить кумуляту; 5) рассчитать моду, медиану, средние, дисперсии, стандартные отклонения и другие параметры.

Типовое задание №3

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) ребенок играет буквами азбуки. Какова вероятность того, что он случайно сложит слово (указано в варианте), т.е. что это слово у него получится при случайном расположении букв в ряд.; 2)рассчитать вероятности вытаскивания на экзамене разного количества знакомых вопрос в билетах, состоящих из 3 вопросов, при условии, что студент выучил только p -долю вопросов; 3) задача на вероятность совместного появления пересекающихся независимых событий; 4) задача на применение теоремы Байеса.

Типовое задание №4

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) как изменится математическое ожидание (среднее) и дисперсия содержания подвижного фосфора (случайная величина x), если в исходном состоянии в пахотном горизонте опытного поля они были равны Q и W соответственно, а по истечении некоторого срока: а) Содержание фосфора уменьшилось на A%, то есть осталось (100-A)%; б) Содержание фосфора стало равным y =x +z, где z = b(c -x); 2) найти вероятности наступления событий для биномиального распределения при объеме выборки n и вероятности p., нахождение среднего и дисперсии, построить графика; 3) Определить вероятность обнаружения x событий в одной пробе (образце, одной учетной рамке), если распределение случайной величины подчиняется закону Пуассона с математическим ожиданием μ=np, где p – вероятность встретить единичное событие, n – объем выборки.

Типовое задание №5

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) построить график интеграла вероятности для стандартного нормального распределения; 2) вычислить для нормально распределенной случайной величины ξ с параметрами 0; 1 вероятность, соответствующую интервалу значений от ξ1 до ξ2; 3) вычислить для нормально распределенной случайной величины x с параметрами μ и б вероятность, соответствующую интервалу значений от x1 до x2, меньше x3 и больше x4; 4) вычислить доверительные границы для нормально распределенной случайной величины ξ с параметрами μ и б, соответствующие доверительным вероятностям 90%, 95% и уровню значимости 0,5.

Типовое задание №6

Заполнить разделы рабочей тетради по теме. 1) Описать свою курсовую работу в терминах статистического эксперимента. 2) Выделить факторы, определяющие элементарные статические испытания. 3) Полная характеристика одного из статических испытаний: общие условия, второстепенные условия, элемент опробования, физическая и статистическая генеральные совокупности, физическая и статистическая выборочные совокупности.

Типовое задание №7

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) найти ошибки репрезентативности для оценок: ошибка среднего, ошибка стандартного отклонения, ошибка коэффициента вариации, ошибка дисперсии; 2) рассчитать доверительные интервалы для среднего (P=95, 99 и 90%); 3) рассчитать гарантированные максимумы и минимумы (P=95, 99 и 90%); 4) рассчитать доверительные границы отдельных значений, условно приняв, что выборочные параметры равны генеральным, а величина распределена нормально (P=95, 99 и 90%); 5) рассчитать доверительные границы для отдельных значений по выборочным оценкам (P=95, 99 и 90%); 6) рассчитать показатели

относительной вероятностной погрешности (Р=95, 99 и 90%); 7) рассчитать интервальную оценку дисперсии (Р=95%); 8) рассчитать интервальную оценку доли всхожести семян для уровня значимости 0,05.

Типовое задание №8

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) рассчитать интервальную оценку дисперсии и стандартного отклонения; 2) проверить возможность аппроксимации распределения экологического свойства нормальным законом с помощью статистики хи-квадрат; 3) проверить возможность аппроксимации распределения одной из характеристик потенциальной засоренности территории законом Пуассона (критерий хи-квадрат);

Типовое задание №9

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) сделать выводы о загрязнении территории по результатам сравнения дисперсий двух обследованных участков с контрольной территорией, о которой известно, что она не загрязнена; 2) сделать выводы о наличии систематической лабораторной ошибки для двух уровней значимости по результатам аттестации стандартного образца; 3) провести оценку значимости различий между генеральными средними двух коррелированных совокупностей, характеризующих измерения, выполненные двумя специалистами в одних и тех же образцах; 4) сравнить генеральные средние двух некоррелированных совокупностей, характеризующих два угодья одного севооборота, если дисперсии однородны.

Типовое задание №10

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) сравнить генеральные средние двух некоррелированных совокупностей, характеризующих два угодья одного севооборота, если дисперсии неоднородны; 2) сравнить встречаемость солонцов (α=0,05) на трех участках по долям площади, занимаемой солонцами; 3) определить повторность для оценки среднего с заданной абсолютной или относительной погрешностью; 4) определить необходимый объем выборок при сравнении двух средних значений.

Типовое задание №11

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) провести однофакторный дисперсионный анализ; 2) оценить степень влияния изучаемого фактора (по формуле Миллса); 3) рассчитать и изобразить графически доверительные интервалы для средних (Р95% и Р99%); 4) оценить существенность различий между средними (по НСР) (Р95% и Р99%); 5) подвести итог дисперсионному анализу.

Типовое задание №12

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) рассчитать коэффициент корреляции и коэффициент детерминации; 2) проверить статистическую значимость коэффициента корреляции (α=0,05) по t-критерию и при помощи z –преобразования; 3) найти повторность для того, чтобы коэффициент корреляции значимо отличался от нуля (α=0,05).

Типовое задание №13

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) провести регрессионный анализ а) y/x б) x/y 2) проверить статистическую значимость коэффициентов (по критерию Стьюдента); 3) построить точечный график (x,y) и отложить обе линии регрессии; 4) Для линейной регрессии y/x оценить значимость регрессии по F-критерию.

Рассчитать ошибки коэффициентов регрессии y/x – и оценить их значимость. Рассчитайте доверительные интервалы для двух значимых коэффициентов регрессии.

Типовое задание №14

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) для связанных (скоррелированных) выборок сравнить средние значения по критерию знаков; 2) проверить гипотезы о равенстве средних с помощью статистики Вилкоксона; 3) провести непараметрический дисперсионный анализ по критерию Краскела-Уоллиса.

Типовое задание №15

Заполнить разделы рабочей тетради по теме. Решить задачи: 1) проверить возможность аппроксимации распределения содержания «физического песка» в пахотном горизонте дерново-подзолистой почвы (%) нормальным законом с помощью статистики Колмогорова – Смирнова; 2) рассчитать коэффициент корреляции Спирмена; 4) рассчитать тетрахорический показатель связи для таблицы и проверить его значимость; 5) Рассчитать коэффициенты ассоциации Дайса.

Типовое задание №16

Заполнить разделы рабочей тетради по теме. Решить задачи на компьютере: 1) познакомиться с программой Statistica 6.0; 2) рассчитать статистические характеристики и исследовать распределения для одного из почвенных свойств (в зависимости от варианта - мощности пахотного горизонта, его влажности, гидрологической кислотности, содержания гумуса, глубины нижней границы горизонта Е), полученных на двух участках дерново-подзолистой (серой лесной, черноземной или др.) почвы, расположенных на разных элементах рельефа;

Типовое задание №17

Заполнить разделы рабочей тетради по теме. Решить задачи на компьютере: на сельскохозяйственных угодьях, расположенных на дерново-подзолистых, серых лесных или черноземных почвах было отобрано примерно по 500 образцов, в которых измерялось содержание гумуса (%) и содержание подвижного фосфора (мг/кг). Нужно проверить: подчиняется ли распределение описанных свойств равномерному и нормальному распределению; проверить, можно ли каждую из изучаемых переменных аппроксимировать нормальным или равномерным распределениями.

Типовое задание №18

Заполнить разделы рабочей тетради по теме. Решить задачи на компьютере: на сельскохозяйственных угодьях, расположенных на агросерых почвах Брянского ополья, измерялось число сорняков на площадке 1 м*1 м по методу рамки. На каждом поле работали студенты сначала под руководством преподавателя, а потом самостоятельно. На каждом поле работало по 2 бригады студентов, которым присвоены условные номера. Нужно проверить: 1) отличаются ли в среднем результаты, полученные студентами, при работе с преподавателем и без преподавателя; 2) одинаково ли работали в среднем две бригады студентов.

Типовое задание №19-20

Заполнить разделы рабочей тетради по теме. Решить задачи на компьютере: провести двухфакторный дисперсионный анализ с целью оценить влияние разных доз удобрений и средств защиты растений на урожайность сельскохозяйственной культуры. Выявить оптимальные дозы удобрений и гербицидов. Сравнить средние по градациям факторов. Проанализировать однофакторные и двухфакторные анализы. Провести непараметрический дисперсионный анализ.

Типовое задание №21-22

Заполнить разделы рабочей тетради по теме. Решить задачи на компьютере: провести множественный линейный регрессионный анализ данных зависимости урожайности от почвенных признаков; определить переменные, влияющие на урожайность; оценить степень влияния каждой из переменных; исследовать качество аппроксимации зависимости; провести анализ остатков.

Типовое задание №23

Заполнить разделы рабочей тетради по теме. Решить задачи на компьютере: 1) провести иерархическую классификацию горизонтов методами одиночной связи и Варда, используя Евклидово расстояние; 2) провести классификацию переменных этими же методами; 3) выполнить два варианта классификации объектов методом k-средних, задав в первом случае 3 класса, во втором - 5 классов.

Типовое задание №24

Заполнить разделы рабочей тетради по теме. Решить задачи на компьютере: провести анализ данных по 6 разрезам дерново-подзолистых почв, представленных горизонтами, методом главных компонент. В образцах было определено содержание гумуса, pH солевой, содержание ила и физической глины, степень ненасыщенности по Гидройц. Рассчитайте собственные числа. Построить график каменной осыпи, проекции переменных и объектов на главные компоненты.

Типовое задание №25

Заполнить разделы рабочей тетради по теме. Решить задачи на компьютере: 1) выполнить дискриминантный анализ совокупности данных о горизонтах дерново-подзолистой почвы, 2) оценить качество классификации; 3) сравнить результаты дискриминантного анализа и метода главных компонент.

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

Примерные вопросы для экзамена 5 семестр

1. Шкалы измерений.
2. Общие и второстепенные условия проведения испытаний.
3. Физическая и статистическая совокупность.
4. Выборка. Техника обеспечения репрезентативности выборки.
5. Статистическая устойчивость. Вероятность. Невозможные, достоверные, несовместимые события.
6. Пересекающиеся события. Независимость событий.
7. Полная группа событий. Теорема Байеса.
8. Закон распределения случайной величины. Константы и параметры распределений.
9. Мода, медиана, среднее. Свойства среднего.
10. Дисперсия, стандартизированное отклонение, коэффициент вариации.
11. Моменты. Коэффициенты асимметрии и эксцесса.
12. Распределение Бернулли. Биномиальное распределение. Распределение Пуассона.
13. Закон нормального распределения.
14. Логнормальное распределение.

15. Оценивание параметров и констант распределений. Качество оценок: несмещенность, эффективность, состоятельность. Робастность.
16. Оценки среднего.
17. Оценки дисперсии и стандартного отклонения. Оценка коэффициента вариации.
18. Ошибки репрезентативности для среднего, дисперсии, стандартного отклонения,.
19. Ошибки коэффициента вариации, коэффициентов асимметрии и эксцесса.
20. Оценка доли и ее ошибка репрезентативности.
21. Интервальные оценки. Распределение Стьюдента. p-значение.
22. Доверительный интервал для среднего. Гарантированные минимум и максимум.
23. Интервал возможных значений признака. Границы типичных значений.
24. Доверительный интервал для доли.
25. Планирование объема выборки для оценки среднего с заданной точностью.
26. Распределение хи-квадрат. Доверительный интервал для дисперсии и стандартного отклонения.
27. Статистическая гипотеза. Статистические критерии. Уровень значимости. Алгоритм проверки статистической гипотезы.
28. Гипотезы о средних значениях: сравнение среднего с константой и сравнение средних двух коррелированных выборок.
29. Распределение Фишера. Проверка гипотезы о равенстве двух дисперсий.
30. Сравнение двух средних: а)дисперсии совокупностей одинаковы и б)дисперсии разные.
31. Сравнение двух долей.
32. Расчет числа повторностей для оценки среднего с заданной абсолютной или относительной погрешностью.
33. Сравнение нескольких средних. Дисперсионный анализ. Условия применимости.
34. Наименьшая существенная разность. Сравнение группы средних.
35. Непараметрические методы сравнения средних. Критерий Вилкоксона, Манна-Уитни, Краскела-Уоллиса.
36. Корреляционный анализ. Коэффициенты корреляции Пирсона и Спирмена. Тетрахорический коэффициент. Коэффициент Дайса.
37. Регрессионный анализ. Оценка значимости регрессии в целом. Оценка значимости регрессионных коэффициентов для простой линейной регрессии.

Примерные вопросы для зачета 6 семестр

1. Опишите любое испытание, в результате которого фиксируется многомерная случайная величина
2. Какими особенностями обладают данные в экологических и почвенных исследованиях?
3. Как можно поступить, если в силу ряда причин часть данных оказалось утеряно:
4. Что является аналогом среднего и дисперсии для многомерной случайной величины?
5. Как формулируется нулевая гипотеза для многомерного дисперсионного анализа?
6. Напишите, как раскладывается общая сумма квадратов для 4-факторного дисперсионного анализа?
7. Приведите пример эксперимента в почвоведении, соответствующего двухфакторному дисперсионному анализу?
8. Условия применимости дисперсионного анализа?

9. Что такое НЗР? Как по-другому оно еще называется и для чего рассчитывается?
10. Какие коэффициенты корреляции можно рассчитать для многомерного случая?
11. Какие условия накладываются на вектора Y и X1, X2…Xp для применения корреляционного и регрессионного анализов?
12. Обобщение одномерных регрессионных моделей на многомерный случай.
13. Приведите пример эксперимента в почвоведении, соответствующего множественной линейной регрессии. Укажите единицы для признаков
14. Как проверяется адекватность предполагаемой модели линейной регрессии?
15. Что такое «гипотеза о влиянии переменной на предсказание Y»?
16. Как рассчитать доверительный интервал для коэффициентов регрессии? Приведите пример
17. Что такое мультиколлинеарность?
18. Для чего проводится стандартизация переменных в регрессионном анализе?
19. Что такое «остатки» и зачем нужен анализ остатков в регрессионном анализе?
20. Приведите пример эксперимента в почвоведении, соответствующего множественной линейной регрессии. Переформулируйте этот пример так, чтобы можно было использовать дисперсионный анализ. Укажите единицы для признаков в обоих случаях
21. Сходство и различие почвенных объектов
22. Какие основные вопросы рассматривает научное направление «методы численной классификации»?
23. Какова обобщенная формула расстояния, введенная по аналогии с физическим расстоянием?
24. Чем отличается Евклидово расстояние от Манхэттенского расстояния?
25. Как можно рассчитать расстояние для сравнения 15 закодированных морфологических описаний разрезов?
26. Что такое меры объединения и какие из них Вы знаете?
27. В чем суть кластерного анализа? Какие методы Вам известны?
28. Какими свойствами должна обладать линейная комбинация X1, X2, …Xp, чтобы стать главной компонентой (МГК)?
29. Ответы на какие вопросы дает метод главных компонент? Приведите примеры
30. В чем задачи дискриминантного анализа? Приведите свой пример.
31. В каких случаях происходит отказ от классификации?

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 соответствующие виды оценочных средств	**2**	**3**	**4**	**5**
Знания Письменный и устный опрос	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения Практические контрольные задания	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) Отчет по практическим заданиям	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

9. **Ресурсное обеспечение:**

- Дмитриев, Евгений Анатольевич. Математическая статистика в почвоведении : [учебник] / Е. А. Дмитриев ; науч. ред. Ю. Н. Благовещенский - Изд. 4-е, доп.- М. : URSS : ЛИБРОКОМ, 2010. Шифр: 3УБ, Д-533
- Самсонова, Вера Петровна. Математическая статистика в почвоведении и экологии : практикум : учеб. пособие / В. П. Самсонова, С. Е. Дядькина, Ю. Л. Мешалкина ; Моск. гос. ун-т им. М. В. Ломоносова. М. : Дашков и К', 2021. Описание Карточка Шифр: 3УБ, С-178
- Мешалкина, Юлия Львовна. Математическая статистика в почвоведении : практикум /Ю.Л. Мешалкина, В.П. Самсонова ; Моск. гос. ун-т им. М.В. Ломоносова М. : МАКС Пресс, 2008. Шифр: 3УБ, М-558
- Пифо, Ханс-Петер. Статистика : для бакалавров по специальностям Агробиология (АБ), Аграрные науки (АН) и Возобновляемые природные ресурсы (ВПР) : в Ун-те Хоэнхайм : 1 семестр (АБ, АН, ВПР), 3 семестр (ВПР) / проф. Ханс-Петер Пифо ; пер. с нем. И. Ф. Кузяковой ; Ин-т с.-х. культур, Отд-ние биоинформатики М. : ВНИИА, 2011 Карточка Шифр: 6, П-357
- Шитиков, Владимир Кириллович. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R / В. К. Шитиков, Г. С. Розенберг

; Рос. акад. наук, Ин-т экологии Волж. Бассейна Тольятти : Кассандра, 2014. Шифр: 3Ф, Ш-648

- Soil&Statistics [Электронный ресурс]. Сайт для обучения математической статистике студентов факультета почвоведения МГУ имени М.В. Ломоносова// Мешалкина Ю.Л., Самсонова В.П. – Режим доступа: https://sites.google.com/site/soilsstatistics/ (дата обращения: 03.08.2022).

- StatSoft [Электронный ресурс]// StatSoft Russia – Режим доступа: http://www.statsoft.ru/ (дата обращения: 03.08.2022).

- The R Project for Statistical Computing [Электронный ресурс] // R Core Team. – Режим доступа: http://www.R-project.org/ (дата обращения: 03.08.2022).

- Информационная система Почвенно-географическая база данных России [Электронный ресурс] // Почвенным дата-центром МГУ имени М.В. Ломоносова. – Режим доступа: https://soil-db.ru/ (дата обращения: 03.08.2022).
- Joint Research Centre
- European Soil Data Centre (Esdac) [Электронный ресурс] // European Commission – Режим доступа: http://esdac.jrc.ec.europa.eu/ (дата обращения: 03.08.2022).

10. **Язык преподавания:**
Русский

11. **Преподаватель (преподаватели):**

Самсонова Вера Петровна, доцент, 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.

Мешалкина Юлия Львовна, доцент, кандидат сельскохозяйственных наук (по специальности 06.01.03 – почвоведение, Специализированный совет К020.25.01 в Почвенном институте имени В.В. Докучаева 15 июня 1989 г. (утв. ВАК), ученое звание доцента по специальности почвоведение (биологические науки) - 1 июля 2005 г. Приказ Высшей Аттестационной Комиссии.

Дядькина Светлана Евгеньевна, научный сотрудник, кандидат биологических наук (по специальности 03.00.27 – почвоведение, Диссовет К.501.001.04., МГУ имени М.В. Ломоносова 19 мая 2004 г. (утв. ВАК)), без ученого звания.

Салпагарова Ирина Азретовна, доцент, кандидат сельскохозяйственных наук (по специальности 06.01.03 - агропочвоведение и агрофизика, Специализированный совет К.020.25.01, Почвенный институт им. В.В. Докучаева 25 октября 1990 г. (утв. ВАК)),

ученое звание доцента по кафедре химии почв – 16 июля 2008 г. Приказ Федеральной службы по надзору в сфере образования и науки.

Бутылкина Марина Александровна, доцент, кандидат биологических наук (по специальности 06.01.14 – агрофизика, Диссовет К053.05.16, МГУ имени М.В. Ломоносова 5 октября 1999 г. (утв. ВАК)), без ученого звания.

Садовникова Надежда Борисовна, старший научный сотрудник, кандидат биологических наук (по специальности 06.01.03 - агропочвоведение, агрофизика, Диссовет Д501.002.13,МГУ имени М.В. Ломоносова 23 декабря 2008 г. (утв. ВАК)), без ученого звания.

Фаустова Екатерина Валерьевна, старший преподаватель, кандидат биологических наук (по специальности 06.01.03 - агрофизика, Диссовет Д501.002.13, МГУ имени М.В. Ломоносова 25 мая 2003г. (утв. ВАК)), без ученого звания.

Липатов Денис Николаевич, старший преподаватель, кандидат биологических наук по специальности 03.00.27 – почвоведение, Диссовет К053.05.16, МГУ имени М.В. Ломоносова 17 октября 2000 г. (утв. ВАК)), без ученого звания.

Рыжова Ирина Михайловна, профессор, доктор биологических наук (по специальности 03.00.27 – почвоведение, Диссовет Д501.001.57, МГУ имени М.В. Ломоносова 2 июня 2006 г. (утв. ВАК)), ученое звание профессор по кафедре Общего почвоведения 28 мая 2013 г. Приказ Высшей Аттестационной Комиссии Минобрнауки РФ.

12. **Разработчики программы:**

Самсонова Вера Петровна, доцент, доктор биологических наук (по 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.

Мешалкина Юлия Львовна, доцент, кандидат сельскохозяйственных наук (по специальности 06.01.03 – почвоведение, Специализированный совет К 020.25.01 в Почвенном институте имени В.В. Докучаева 15 июня 1989 г. (утв. ВАК), ученое звание доцента по специальности почвоведение (биологические науки) - 1 июля 2005 г. Приказ Высшей Аттестационной Комиссии.

13. **Краткая аннотация дисциплины:**

Целью дисциплины является освоение теоретических и практических основ обучающихся в области применения математической статистики в почвоведении и экологии, владение системой фундаментальных научных понятий, методологией и методами современной прикладной статистики, готовность проводить научные исследования в области анализа данных, способность применять и разрабатывать новые инновационные технологии,

приобретение ими практических навыков и компетенций в сфере применения математической статистики в почвоведении и экологии.

Тип образовательного стандарта - интегрированный магистр МГУ, учебный план бакалавра; направление подготовки: 06.03.02 «Почвоведение». Дисциплина относится к базовой части к разделу – общенаучный. Она читается на 3 курсе (5 и 6 семестры). Общая трудоемкость 144 в академических часах, 4 в зачетных единицах. Студенты, изучают следующие разделы: случайная величина как модель почвенного покрова, классические методы анализа единичной выборки, анализ группы выборок, непараметрические методы анализа, многомерные случайные величины, исследование зависимостей в случае многомерных данных, методы численной классификации. Форма промежуточной аттестации: экзамен и зачет. По окончании курса студенты должны иметь представление о генезисе данных о почве и почвенном покрове, знать основные понятия прикладной статистики, ориентироваться в методах обработки единичных выборок и групп выборок, уметь пользоваться статистическими пакетами.