


ГОСУДАРСТВЕННОЕ НАУЧНОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
«АКАДЕМИЯ НАУК РЕСПУБЛИКИ ТАТАРСТАН»
ИНСТИТУТ ПРОБЛЕМ ЭКОЛОГИИ И НЕДРОПОЛЬЗОВАНИЯ

СОГЛАСОВАНО

Вице-президент АН РТ




А. А. Тимерханов
«10» июня 2024 г.

УТВЕРЖДЕНО

Руководитель (директор) Института
проблем экологии и недропользования
АН РТ



Р. Р. Шагидуллин
«2» июня 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

2.1.5. Статистические методы анализа данных

Уровень: подготовка научных и научно-педагогических кадров в аспирантуре

Научная специальность: 1.5.15 Экология

Профиль: по отраслям

Нормативный срок освоения программы: 4 года


Форма обучения: очная

Казань 2024

Разработчик:

С.н.с. лаборатории биомониторинга
ИПЭН АН РТ, к.б.н.  Р.А.Суходольская

Рабочая программа дисциплины одобрена Ученым советом Института проблем экологии и недропользования АН РТ, протокол №2/24 от 2.07.24 г.

Зам.директора по научной работе
Института проблем экологии
и недропользования АН РТ, д.г.н.  Д.В.Иванов

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цель: обучение аспирантов современным методам статистического анализа данных с использованием компьютерных программ - пакетов статистической обработки данных (Microsoft Excel, STATISTICA и др.), оформления результатов в виде и табличного и графического материалов.

Задачи:

- обладать теоретическими основами биометрии;
- знать свойства и характеристики вариационных рядов, критерии выбора методов статистической обработки, оценки достоверности статистических величин;
- уметь выбирать осмысленно статистические методы и правильно интерпретировать результаты расчетов;
- ориентироваться в справочной литературе, статистических таблицах и программном обеспечении;
- обладать навыками оформления результатов статистической обработки в виде и табличного и графического материалов;
- приобрести навыки расчетов статистических показателей с использованием персональных компьютеров.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП АСПИРАНТУРЫ

Дисциплина «Статистические методы анализа данных» входит в Блок 2 «Образовательный компонент», программы и читается на 2 курсе по научной специальности «1.5.15 Экология (по отраслям)».

3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ

Дисциплина «Статистические методы анализа данных» направлена на формирование у аспирантов следующих знаний, умений, навыков:

- способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях;
- готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и научно-образовательных задач;
- способностью планировать и решать задачи собственного профессионального и личностного развития;
- способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий;
- способность анализировать и интерпретировать полученные результаты исследований с использованием современных методов обработки и интерпретации информации, в том числе с использованием современных методов статистического анализа.

4. СТРУКТУРА, ОБЪЕМ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины составляет 3 з.е. (108 часов). Время проведения 1 семестр 2 года обучения (3 семестр).

Таблица 1

Структура дисциплины, виды и объем учебной работы

№ п/п	Наименование раздела дисциплины	Виды занятий и трудоемкость в часах					Всего
		Л	С	П	ЛЗ	СР	
1	Тема 1. Вариационный ряд и его основные характеристики	2				2	4
2	Тема 2. Законы и параметры распределения. Нормальное распределение и его признаки	4				2	6
3	Тема 3. Непараметрические критерии	4				2	6
4	Тема 4. Краткая характеристика программных пакетов Microsoft Excel и STATISTICA	4		2			6
5	Тема 5. Статистическая обработка данных в Microsoft Excel и Statistica	12		22		30	64
6	Тема 6. Методы графического анализа данных и представление табличного и графического материала	4		4		2	10
7	Подготовка к зачету					10	10
	Контроль						2
	Итого:	30		28		48	108

Примечание: Л – лекции, С – семинары, П – практические занятия, ЛЗ – лабораторные занятия, СР – самостоятельная работа.

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 2

Содержание разделов дисциплины

№ п/п	Наименование раздела	Содержание раздела
Раздел 1. Методологические основы организации научной деятельности		
1	Тема 1. Вариационный ряд и его основные характеристики	Генеральная совокупность. Выборка и ее объем выборки. Типы данных: интервальные, классификационные (качественные), альтернативные, порядковые.
2	Тема 2. Законы и параметры распределения. Нормальное распределение и его признаки	Нормальное распределение и его признаки. Параметрические критерии: t-критерий Стьюдента, а для оценки дисперсии - F-критерий Фишера. Основные критерии и параметры вариационного рядов нормального распределения: средняя арифметическая (M), среднее квадратическое отклонение (стандартное отклонение или σ), показатели вариации (CV), дисперсия, стандартная ошибка (m),

		достоверность средней арифметического, точность определения средней, асимметрия, эксцесс. ANOVA. Проверка статистических гипотез.
3	Тема 3. Непараметрические критерии	Непараметрические критерии: T -критерий Уайта, X -критерий Ван-дер-Вардена и U -критерий Уилкоксона (Wilcoxon-test), критерий знаков z , медиана, ранговая корреляция Спирмена и т.д.
4	Тема 4. Краткая характеристика программных пакетов Microsoft Excel и STATISTICA	Краткая характеристика программных пакетов Microsoft Excel и STATISTICA. Модули в программной среде STATISTICA. Основные статистики и таблицы. Вычисление среднеарифметического значения, стандартного отклонения, ошибки среднего, медианы, моды, объема выборки, минимума, максимума, коэффициентов асимметрии и эксцесса. Сравнения переменных с целью оценки достоверности различия между ними.
5	Тема 5. Статистическая обработка данных в Microsoft Excel и Statistica	Вычислительный модуль и модуль диаграмм Microsoft Excel. Электронные таблицы на листе Microsoft Excel и обработка данных с помощью Мастера функций. Пакет Анализ данных, критерии параметрического распределения, статистические анализы, функции распределения, корреляционный анализ и др. в программе Microsoft Excel. Виды графики в Microsoft Excel, оформляемые с помощью Мастера диаграмм. Вставка линии тренда и оценка достоверности аппроксимации линии тренда. Статистическая обработка данных в модуле Nonparametric Distribution. Расчеты непараметрических критериев: медианы (вместо среднеарифметического), X -критерия Ван-дер-Вардена и U -критерия Уилкоксона (Wilcoxon-test), критерия знаков z , ранговой корреляции Спирмена. Методы многомерной статистики. Анализ главных компонент, кластерный и дискриминантный анализы.
6	Тема 6. Методы графического анализа данных и представление табличного и графического материала	Методы графического анализа зоологических данных. Основы компьютерной подготовки научных работ (диссертаций), редактирование и оформление текстовых файлов и графических изображений.

6. ПЕРЕЧЕНЬ ЛЕКЦИЙ, СЕМИНАРСКИХ, ПРАКТИЧЕСКИХ ЗАНЯТИЙ, ЛАБОРАТОРНЫХ И САМОСТОЯТЕЛЬНЫХ РАБОТ

Таблица 3

Перечень занятий и формы контроля

№ п/п	Наименование раздела	Вид занятия	Тема занятия (самостоятельной работы)	Форма текущего и промежуточного контроля
Раздел 1. Методологические основы организации научной деятельности				

1	Тема 1. Вариационный ряд и его основные характеристики	Л	1. Генеральная совокупность. Выборка и ее объем выборки. 2. Предельно допустимые приемы при обработке данных.	УО
		СР	Типы данных: интервальные, классификационные (качественные), альтернативные, порядковые.	КЛ, УО
2	Тема 2. Законы и параметры распределения. Нормальное распределение и его признаки	Л	Нормальное распределение и его признаки. Параметрические критерии: t -критерий Стьюдента, а для оценки дисперсии - F -критерий Фишера.	УО
		Л	Основные критерии и параметры вариационного рядов нормального распределения: средняя арифметическая (M), среднее квадратическое отклонение (стандартное отклонение или σ), показатели вариации (CV), дисперсия, стандартная ошибка (m), достоверность средней арифметического, точность определения средней, асимметрия, эксцесс.	Д, УО, ГД
		СР	Проверка статистических гипотез.	КЛ, УО
3	Тема 3. Непараметрические критерии	Л	1. Непараметрические критерии: T -критерий Уайта, X -критерий Ван-дер-Вардена 2. U -критерий Уилкоксона (Wilkokson-test), критерий знаков z ,	УО
		СР	Ранговая корреляция Спирмена	КЛ, УО
4	Тема 4. Краткая характеристика программных пакетов Microsoft Excel и STATISTICA	Л	Краткая характеристика программных пакетов Microsoft Excel и STATISTICA.	УО
		ПЗ	Модули в программной среде STATISTICA. Основные статистики и таблицы. Вычисление среднеарифметического значения, стандартного отклонения, ошибки среднего, медианы, моды, объема выборки, минимума, максимума, коэффициентов асимметрии и эксцесса. Сравнения переменных с целью оценки достоверности различия между ними.	КЛ
5		Л	1. Вычислительный модуль и модуль диаграмм Microsoft Excel. 2. Электронные таблицы на листе Microsoft Excel и обработка данных с помощью 3. Мастера функций. Пакет 4. Анализ данных, критерии параметрического распределения, статистические анализы, функции распределения, корреляционный анализ и др. в программе Microsoft Excel.	УО, ГД

			5. Виды графики в Microsoft Excel, оформляемые с помощью Мастера диаграмм. 6. Вставка линии тренда и оценка достоверности аппроксимации линии тренда.	
		ПЗ	1. Статистическая обработка данных в модуле Nonparametric Distribution. 2. Расчеты непараметрических критериев: медианы (вместо среднеарифметического). 3. X-критерия Ван-дер-Вардена и U-критерия Уилкоксона (Wilkinson-test), критерия знаков z, ранговой корреляции Спирмена. 4. Методы многомерной статистики. 5. Анализ главных компонент. 6. Кластерный анализ. 7. Дискриминантный анализ 8. ANOVA 9. MANOVA 10. Бут-стреп. 11. Факторный анализ.	П
6	Тема 6. Методы графического анализа данных и представление табличного и графического материала	Л	1. Методы графического анализа зоологических данных. 2. Основы компьютерной подготовки научных работ (диссертаций), редактирование и оформление текстовых файлов и графических изображений.	П
		ПЗ, СР	1. Графически оформить данные матрицы 2. Форматировать таблицу матричных данных	П
			Представить рисунок по матричным данным	
Итоговый контроль				Зачет

Виды занятий: Л – лекции, С – семинары, П – практические занятия, ЛЗ - лабораторные занятия, СР – самостоятельная работа.

Формы текущего контроля: УО - устный опрос (собеседование), Р - реферат, П - проект, Д - доклад, КЛ - конспект лекции, ГД - групповая дискуссия и др.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

Таблица 4

Карта обеспечения учебно-методической литературой

№ п/п	Автор, название, место издания, издательство, год издания учебной и учебно-методической литературы	Количество экз.	Число аспирантов, одновременно изучающих дисциплину
-------	--	-----------------	---

Основная литература

1.	Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL. М.: Форум: Инфра-М, 2014. 464 с. (библиотека)	1	1
2.	Гринин А.С. Математическое моделирование в экологии. М.: ЮНИТИ-ДАНА, 2003. 269 с. (библиотека)	1	
3.	Пузаченко Ю.Г. Математические методы в экологических и географических исследованиях. М.: ИЦ "Академия", 2004. 416 с. (Библиотека)	1	
4.	Новиков Д.А. Статистические методы в медико-биологическом эксперименте (типовые случаи) [Электронный ресурс]: учебное пособие/ Новиков Д.А., Новочадов В.В.— Электрон. текстовые данные.— Вологод: Издательство ВолГМУ, 2005.— 84 с.— Режим доступа: http://www.iprbookshop.ru/8502 . — ЭБС «IPRbooks», по паролю		В свободном доступе с компьютеров ИПЭН АН РТ и по паролю при удаленном доступе
5.	Карташов Г.Д. Многомерный статистический анализ [Электронный ресурс]: методические указания к выполнению курсовой работы/ Карташов Г.Д., Тимонин В.И., Будовская Л.М.— Электрон. текстовые данные.— М.: Московский государственный технический университет имени Н.Э. Баумана, 2007.— 48 с.— Режим доступа: http://www.iprbookshop.ru/31083 . — ЭБС «IPRbooks», по паролю		
6.	Введение в статистический анализ медицинских данных [Электронный ресурс] : учебное пособие для аспирантов / Д.Н. Бегун [и др.]. — Электрон. текстовые данные. — Оренбург: Оренбургская государственная медицинская академия, 2014. — 118 с. — 2227-8397. — Режим доступа: http://www.iprbookshop.ru/54283.html		
7.	Пашкевич О.И. Статистическая обработка эмпирических данных в системе STATISTICA [Электронный ресурс] : учебно-методическое пособие / О.И. Пашкевич. — Электрон. текстовые данные. — Минск: Республиканский институт профессионального образования (РИПО), 2014. — 148 с. — 978-985-503-385-2. — Режим доступа: http://www.iprbookshop.ru/67607.html		
8.	Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход [Электронный ресурс] : монография / Б.Ю. Лемешко [и др.]. — Электрон. текстовые данные. — Новосибирск: Новосибирский		

	государственный технический университет, 2011. — 888 с. — 978-5-7782-1590-0. — Режим доступа: http://www.iprbookshop.ru/47719.html		
--	---	--	--

Таблица 5

Перечень печатных, технических и электронных средств обучения

№ п/п	Наименование	Вид	Форма доступа
1	Основы статистики https://stepik.org/course/76/	Сайт	Свободный доступ
2	Анализ данных https://ru.coursera.org/specializations/analiz-dannykh	Сайт	Свободный доступ
3	Анализ данных в R https://stepik.org/course/129/	Сайт	Свободный доступ
4	Статистика для гуманитариев https://openedu.ru/course/tgu/Stat/	Сайт	Свободный доступ

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Таблица 6

Обеспеченность помещениями для аудиторных занятий и мультимедийного оборудования

№ п/п	Наименование дисциплин в соответствии с учебным планом, вид занятий	Наименование специализированных аудиторий, кабинетов, лабораторий и пр. с перечнем основного оборудования	Форма владения, пользования (собственность, оперативное управление, аренда и т.п.)

	Статистические методы анализа данных	<p><u>1. АКТОВЫЙ ЗАЛ (90,7 кв.м):</u> Радиосистема WMS 40 mini dual – 2 шт.; Радиомикрофон – 4 шт. Микрофон – 2 шт. Микшер Yamaha MG123cx/c – 1 шт.; Ноутбук Samsung NP-RF711 – 1 шт.; Проектор Nec v300x 3D Ready (V300x6) – 1 шт.; Экран настенный Classic Norma 244x244 (W236x236/1 MW-L4/W) – 1 шт.; Стол переговорный – 6 шт.; Стол компьютерный угловой – 1 шт.; Кресло «Лотос» (черное) – 21 шт.; Стул CM-7 (кожзам) – 12 шт.; Кресло для залов – 30 шт.</p> <p><u>2. Библиотека (30,5 кв.м):</u> Стол – 2 шт.; Стулья – 6 шт.; МФУ Kyocera Taskalfa 220 – 1 шт.; Персональный компьютер – 2 шт.</p>	Оперативное управление
--	--------------------------------------	---	------------------------

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Для реализации программы при изучении учебной дисциплины «Статистические методы анализа данных» используются активные формы обучения: лекции, вариативный опрос, дискуссии, устный опрос. Предусмотрены практические занятия.

9. ОЦЕНОЧНЫЕ СРЕДСТВА

Фонд оценочных средств для проведения текущего контроля (ФОС ТК) является составной частью РП дисциплины.

Типовые оценочные средства для текущего контроля

Вопросы для устного опроса

Тема 1. Вариационный ряд и его основные характеристики.

1. Что означает ось X при построении вариационного ряда?
2. Насколько может выражаться эксцесс?
3. Параметры асимметрии: сущность.

Тема 2. Законы и параметры распределения.

1. Зачем вычисляется t-критерий Стьюдента?
2. Чем отличается дисперсия от среднеквадратичного отклонения?
3. Чем отличаются способы вычисления эксцесса и асимметрии?

Тема 3. Непараметрические критерии.

1. Чем непараметрические критерии отличаются от параметрических?
2. Почему ранговая корреляция Спирмена не вычисляется при обработке континуальных данных?
3. Чем примечателен критерий знаков?

Тема 4. Краткая характеристика программных пакетов Microsoft Excel и STATISTICA.

1. Что такое модуль?
2. Чем среднеарифметическое отличается от медианы?
3. Назовите способы сравнения переменных.

Тема 5. Статистическая обработка данных в Microsoft Excel и Statistica.

1. Почему не стоит применять Мастер функций в пакете Excel?
2. Какими параметрами характеризуются графики в пакете Excel?
3. Назовите типы линии трендов.

Тема 6. Методы графического анализа данных и представление табличного и графического материала.

1. В чем состоит преимущество графического оформления данных?

Оценивание ответов на устный опрос проводится по системе зачтено/не зачтено в соответствии со следующими критериями:

Критерии оценивания

Оценка	Критерии
зачтено	Полнота ответа на поставленный вопрос, умение использовать термины, приводить примеры, делать выводы и анализировать конкретные ситуации
не зачтено	При ответе аспиранта выявились существенные пробелы в знаниях основных положений дисциплины, неумение с помощью преподавателя получить правильное решение конкретной задачи

Практические задания

Практическое задание №1

Оформите собственные данные в виде файла в пакете Excel

Практическое задание № 2.

Оформите собственные данные в виде файла в пакете Statistica

Практическое задание №3.

Рассчитайте основные статистические характеристики по заданной матрице в пакете Excel.

Практическое задание №4.

Рассчитайте основные статистические характеристики по заданной матрице в пакете Statistica.

Практическое задание №5

Проведите регрессионный анализ в пакете Excel.

Практическое задание №6.

Проведите регрессионный анализ в пакете Statistica.

Практическое задание №7.

Проведите корреляционный анализ в пакете Excel.

Практическое задание №8.

Проведите корреляционный анализ в пакете Statistica.

Практическое задание №9.

Проведите анализ главных компонент в пакете Statistica.

Практическое задание №10.

Проведите дискриминантный анализ в пакете Statistica.

Практическое задание №11.

Проведите бутстрепв пакете Statistica.

Практическое задание №12.

Оформите графически заданную матрицу в пакете Excel.

Практическое задание №13.

Оформите графически заданную матрицу в пакете Statistica.

Оценивание результатов выполнения практических заданий проводится по системе зачтено/не зачтено в соответствии со следующими критериями:

Оценка	Критерии
зачтено	Полнота ответа на поставленный вопрос, умение использовать термины, приводить примеры, делать выводы и анализировать конкретные ситуации
не зачтено	При ответе аспиранта выявились существенные пробелы в знаниях основных положений дисциплины, неумение с помощью преподавателя получить правильное решение конкретной задачи

Оценочные средства для промежуточной аттестации

Фонд оценочных средств для проведения промежуточной аттестации (ФОС ПА) является составной частью РП дисциплины, разработан в виде отдельного документа, в соответствии с положением о ФОС ПА.

Примерные вопросы на зачете:

1. Основные характеристики вариационного ряда.
2. Параметрические критерии оценки данных.
3. Непараметрические критерии оценки данных.
4. Краткая характеристика модулей пакета Statistica.
5. Краткая характеристика модулей пакета Excel.
6. Основные приемы многомерного анализа и области его применения.
7. Выполнить вычисление коэффициента Стьюдента по заданным данным.
8. Выполнить описательную статистику по заданным данным.
9. Выполнить кластерный анализ по заданным данным.
10. Выполнить анализ главных компонент по заданным данным.
11. Выполнить дискриминантный анализ по заданным данным.
12. Графическое представление данных в пакете Statistica.
13. Характеристика модуля анализа главных компонент.
14. Характеристика модуля дискриминантного анализа.
15. Особенности формирования матриц при проведении многомерного анализа.

Критерии оценки промежуточной аттестации

Оценка	Критерии
зачтено	Аспирант показал знания основных положений дисциплины, умение решать конкретные практические задачи, предусмотренные рабочей программой, ориентироваться в рекомендованной справочной литературе

Не зачтено	Аспирант не выполнил основные требования к аттестации в текущем семестре (в т.ч. по публикационной активности и апробации НИД).
------------	---