

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное учреждение науки
**ЛИМНОЛОГИЧЕСКИЙ ИНСТИТУТ СИБИРСКОГО ОТДЕЛЕНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ЛИН СО РАН)**



УТВЕРЖДАЮ

Директор  А.П. Федотов
«16»  2022 г.

Рабочая программа дисциплины (модуля)

Индекс дисциплины по УП: 3.2(Э)

Наименование дисциплины (модуля):

Методы статистической обработки данных

Подготовка научных и научно-педагогических кадров в аспирантуре

Иркутск, 2022 г.

Содержание

1.	Цель и задачи дисциплины (модуля)	3
2.	Место дисциплины (модуля) в структуре ОПОП	3
3.	Требования к результатам освоения дисциплины (модуля)	3
4.	Объем дисциплины (модуля) и виды учебной работы	4
5.	Содержание дисциплины (модуля)	4
	5.1 Содержание разделов и тем дисциплины (модуля)	4
	5.2 Разделы и темы дисциплин и виды занятий	6
6	Темы практических занятий	7
7	Учебно-методическое и информационное обеспечение дисциплины (модуля)	8
	7.1. Литература	8
	7.2. Программное обеспечение	10
	7.3. Базы данных, информационно-справочные и поисковые системы	10
8	Материально-техническое обеспечение дисциплины (модуля)	11
9	Образовательные технологии	11
10	Кадровое обеспечение	11
11	Оценочные средства	11
	ПРИЛОЖЕНИЕ А	12
	ЛИСТ ОБНОВЛЕНИЙ	15

1 Цели и задачи дисциплины (модуля):

Целью курса является ознакомление с современными направлениями исследований в различных областях науки, использующими методы теории вероятностей, математической статистики и математического моделирования, изучение методов систематизации, классификации и извлечения информации из больших объемов первичных экспериментальных данных, обзор примеров анализа данных, полученных в ходе комплексных исследований различных водоемов.

Задачи дисциплины:

- Изучение теоретических основ методов используемых в математической статистики и математическом моделировании.
- Знакомство с методами описательной статистики, методами оценки параметров распределения и тестирования статистических гипотез, методами многомерной статистики.
- Изучения методов математического моделирования и теории динамических систем

2 Место дисциплины (модуля) в структуре ОПОП:

Программа дисциплины (модуля) «Методы статистической обработки данных» (3.2(Э)) относится к элективным дисциплинам образовательного компонента программы подготовки научных и научно-педагогических кадров в аспирантуре по всем специальностям.

Курс предполагает наличие базовых знаний, полученных по основным программам вуза, включая высшую математику.

3 Требования к результатам освоения дисциплины (модуля):

В результате освоения дисциплины аспирант должен:

Знать:

- основные логические методы и приемы научного исследования, методологические теории и принципы современной науки;
- принципы организации научных исследований с применением методов математического моделирования и математической статистики;
- теоретические основы описательной математической статистики;
- принципы оценки параметров распределений с использованием бутстреп метода максимального правдоподобия и байесовского оценивания;
- теоретические основы методов тестирования статических гипотез;
- теоретические основы корреляционного и регрессионного анализа;
- основы методов многомерной статистики;
- основы методов математического и компьютерного моделирования различных биологических процессов;

Уметь:

- планировать экспериментальные и полевые работы с учетом прогнозов размеров выборок оцениваемых данных;
- оценивать параметры распределений в выборках статистических величин;

- проводить тестирование статистических гипотез, параметрическими и непараметрическими методами;
- проводить графическую визуализацию результатов статического анализа;
- Интерпретировать различные варианты представления результатов статического анализа.

Владеть:

- теоретическими основами математической статистики и математического моделирования;
- современными программными средствами для статических расчетов и математического моделирования применяемых в научных исследованиях;
- Методами визуализации векторной научной графики;

4 Объем дисциплины (модуля) и виды учебной работы

Вид учебной работы		Всего часов / зачетных единиц	Курс
			2
Аудиторные занятия (всего)		112/3,11	112/3,11
В том числе:			
Лекции		50/1,38	50/1,38
Практические и лабораторные занятия		62/1,72	62/1,72
Самостоятельная работа (всего)		28/0,78	28/0,78
Подготовка к текущему контролю и промежуточной аттестации		28/0,78	28/0,78
Вид промежуточной аттестации (зачет)		4/0,11	4/0,11
Общая трудоемкость	часы	144	144
	зачетные единицы	4	4

5. Содержание дисциплины (модуля)

5.1 Содержание разделов и тем дисциплины (модуля):

Тема 1 Численные методы математики. Понятие функция, численные методы исследования поведения функций. Линейные и не линейные уравнение и их системы. Численное решение систем линейных и не линейных уравнений. Дифференциальные уравнения и их системы. Численное интегрирование дифференциальных уравнений.

Тема 2 Основы теории динамических систем и математического моделирования. Понятие динамическая система. Описание динамических систем с помощью дифференциальных уравнений. Понятие стационарных состояний в динамических системах. Методы исследования поведения сложных динамических систем.

Тема 3 Основы теории вероятностей. Понятие вероятности. Случайные величины. Понятие о законе распределения случайных величин, функции распределения

и функции плотности вероятности. Параметры распределений и моменты распределений. Основные законы распределений случайных величин.

Тема 4 Методы оценки параметров распределений статических величин. Понятие выборочной совокупности и генеральной совокупности. Оцениваемое значение случайной величины, абсолютные и относительные ошибки оценивания. Оценка параметров для некоторых распространенных видов распределений. Бутстреп метод для оценки параметров распределений.

Тема 5 Проверка статистических гипотез. Понятие статическая гипотеза. Нулевые и альтернативные статистические гипотезы. Методы проверки и тестирования законов распределения. Методы тестирования гипотез при сопоставлении параметров распределений непрерывных величин. Методы тестирования гипотез при сопоставлении параметров распределений дискретных величин и распределений качественных признаков. Бутстреп как метод тестирования статистических гипотез.

Тема 6 Дисперсионный и корреляционный анализ. Понятие дисперсионного анализа. Однофакторный и многофакторный дисперсионный анализ. Параметрические и не параметрические методы дисперсионного анализа. Перестановочный (бутстреп метод) тест в дисперсионном анализе. Дисперсионный анализ для выборок качественных признаков. Понятие корреляционного анализа. Линейный корреляционный анализ. Различные способы расчета коэффициентов линейной корреляции. Тестирование достоверности коэффициентов линейной корреляции.

Тема 7 Регрессионный анализ. Понятие регрессионного анализа. Различение виды регрессионных функция (регрессионных моделей). Регрессионный анализ методом наименьших квадратов. Тестирование достоверностей результатов регрессионного анализа. Доверительные интервалы для оцененных параметров регрессионных моделей.

Тема 8 Методы многомерной статистики. Многомерные статистические данные. Манипуляция многомерными данными. Различные способы нормировки исследуемых величин. Способы расчета дистанций, различные метрики дистанций, выбор способа расчета дистанций. Кластерный анализ, способы кластеризации. Алгоритмы выбора количества классов. Метод многомерного шкалирования. Анализ методом главных компонент.

Тема 9 Метод максимального правдоподобия и Байесовский статистический анализ. Введение в метод максимального правдоподобия. Оценка параметров распределений методом максимального правдоподобия. Тест отношения показателей правдоподобия. Показатель правдоподобия и информационные критерии при тестировании статистических гипотез. Введение в Байесовское статистическое оценивание. Оценка параметров распределений Байесовским статистическим методом. Тестирование статистических гипотез байесовским статистическим методом.

Тема 10 Применение математических методов в биологических исследованиях. Математические модели динамики популяций. Вычислительная экология. Методы многомерной статистики в экологических исследованиях и химических исследованиях. Анализ взаимосвязи параметров среды с видовым разнообразием.

5.2 Разделы и темы дисциплины (модуля) и виды занятий

№	Темы, разделы	Всего часов	Виды занятий в часах		
			Лекции (зачет)	Практические занятия	Самостоятельная работа
1.	Основы математического моделирования, численные методы математики.	16	6	8	2
2	Основы теории динамических систем	8	4	2	2
3.	Основы теории вероятностей.	6	4	-	2
4.	Методы оценки параметров распределений статистических величин	17	6	5	6
5.	Проверка статистических гипотез.	12	5	5	2
6.	Дисперсионный и корреляционный анализ.	17	5	10	2
7.	Регрессионный анализ.	13	4	7	2
8.	Методы многомерной статистики.	17	6	7	4
9.	Метод максимального правдоподобия и Байесовский статистический анализ.	13	5	6	2
10	Применение математических методов в биологических исследованиях, химических науках и науках об атмосфере и климате	21	5	12	4
	Промежуточная аттестация (зачет)	4	4	-	-
ВСЕГО (часы)		144	54	62	28

6. Темы практических занятий

№ п/п	№ раздела и темы дисциплины	Наименование практических работ	Трудоемкость (часы)	Оценочные средства
1	1	Численные методы математического анализа	4	Контрольные вопросы и задания
2	1	Численное решение линейных и нелинейных уравнений	4	Контрольные вопросы и задания
3	2	Численное решение дифференциальных уравнений	2	Контрольные вопросы и задания
4	4	Параметрические и непараметрические методы оценки параметров распределений	5	Контрольные вопросы и задания
5	5	Проверка статистических гипотез с помощью стандартных статистических методов	5	Контрольные вопросы и задания
6	5	Проверка статистических гипотез с помощью бутстреп метода	10	Контрольные вопросы и задания
7	6	Дисперсионный анализ на с использованием различных методов	5	Контрольные вопросы и задания
8	6	Корреляционный анализ с использованием различных методов	5	Контрольные вопросы и задания
9	7	Регрессионный анализ с помощью метода наименьших квадратов	7	Контрольные вопросы и задания
10	8	Кластерный анализ	1	Контрольные вопросы и задания
11	8	Многомерное шкалирование	2	Контрольные вопросы и задания
12	8	Анализ методом главных компонент	4	Контрольные вопросы и задания
13	9	Анализ с использованием метода максимального правдоподобия	2	Контрольные вопросы и задания
14	9	Байесовский анализ с эле-	4	Контрольные во-

		ментами Монте-Карло моделирования		просы и задания
15	10	Математические модели динамики	4	Контрольные вопросы и задания
16	10	Вычислительная и математическая химия	8	Контрольные вопросы и задания

7 Учебно-методическое и информационное обеспечение дисциплины (модуля):

7.1 Литература

Основная:

1 **Балдин, К.В.** Теория вероятностей и математическая статистика [Электронный ресурс]: учебник / К.В. Балдин, В.Н. Башлыков, А.В. Рукосуев. — Электрон.текстовые данные. — М.: Дашков и К, 2016. — 472 с. — Режим доступа: <http://www.iprbookshop.ru/62453.html>

2 **Губарь, Ю.В.** Введение в математическое моделирование [Электронный ресурс]: учебное пособие / Ю.В. Губарь. — 2-е изд. — Электрон.текстовые данные. — М.: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. — 178 с. — Режим доступа: <http://www.iprbookshop.ru/73662.html>

3 **Гурьянова, И.Э.** Теория вероятностей и математическая статистика. Теория вероятностей. Краткий курс с примерами [Электронный ресурс]: учебное пособие / И.Э. Гурьянова, Е.В. Левашкина. — Электрон.текстовые данные. — М.: Издательский Дом МИСиС, 2016. — 106 с. — Режим доступа: <http://www.iprbookshop.ru/64202.html>

4 **Климов, Г.П.** Теория вероятностей и математическая статистика [Электронный ресурс]: учебник / Г.П. Климов. — Электрон.текстовые данные. — М.: Московский государственный университет имени М.В. Ломоносова, 2011. — 368 с. — Режим доступа: <http://www.iprbookshop.ru/13115.html>

5 **Кобзарь, А.И.** Прикладная математическая статистика. Для инженеров и научных работников [Текст]: учеб.пособие / А.И. Кобзарь. — Электрон.дан. — Москва: Физматлит, 2012. — 816 с. — Режим доступа: библиотечный фонд ЛИИ СО РАН.

Дополнительная:

а) Книжные издания:

6 **Васильев, А.Н.** Python на примерах [Электронный ресурс]: практический курс по программированию / А.Н. Васильев. — 2-е изд. — Электрон.текстовые данные. — СПб.: Наука и Техника, 2017. — 432 с. — Режим доступа: <http://www.iprbookshop.ru/73043.html>

7 **Зиновьева, Е.А.** Компьютерный дизайн. Векторная графика [Электронный ресурс]: учебно-методическое пособие / Е.А. Зиновьева. — Электрон.текстовые данные. — Екатеринбург: Уральский федеральный университет, ЭБС АСВ, 2016. — 116 с. — Режим доступа: <http://www.iprbookshop.ru/68251.html>

8 Карасев, В.А. Теория вероятностей и математическая статистика. Математическая статистика [Электронный ресурс]: практикум / В.А. Карасев, Г.Д. Лёвшина. — Электрон.текстовые данные. — М.: Издательский Дом МИСиС, 2016. — 120 с. — Режим доступа: <http://www.iprbookshop.ru/64203.html>

9 Кауфман, В.Ш. Языки программирования. Концепции и принципы [Электронный ресурс]: / В.Ш. Кауфман. — Электрон.текстовые данные. — Саратов: Профобразование, 2017. — 464 с. — Режим доступа: <http://www.iprbookshop.ru/64055.html>

10 Куренков, А.М. Статистика [Электронный ресурс]: учебник / А.М. Куренков. — Электрон.текстовые данные. — М.: Перспектива, 2012. — 770 с. — Режим доступа: <http://www.iprbookshop.ru/12751.html>

11 Лагутин, М.Б. Наглядная и математическая статистика [Текст]: учеб.пособие / М.Б. Лагутин — 2-е изд., испр. — Москва: Издательство БИНОМ. "Лаборатория знаний", 2009. — 472 с. — Режим доступа: библиотечный фонд ЛИИ СО РАН.

12 Леск, А. М. Введение в биоинформатику = introduction to bioinformatics [Текст]: учебное пособие / А. М. Леск; пер. с англ. под ред. А. А. Миронов и В. К. Швядас. - М.: БИНОМ, Лаборатория знаний, 2009. - 318 с. — Режим доступа: библиотечный фонд ЛИИ СО РАН.

13 Лисицин, Д.В. Методы построения регрессионных моделей [Электронный ресурс]: учебное пособие / Д.В. Лисицин. — Электрон.текстовые данные. — Новосибирск: Новосибирский государственный технический университет, 2011. — 77 с. — Режим доступа: <http://www.iprbookshop.ru/45390.html>

14 Матальцкий, М.А. Теория вероятностей, математическая статистика и случайные процессы [Электронный ресурс]: учебное пособие / М.А. Матальцкий, Г.А. Хацкевич. — Электрон.текстовые данные. — Минск: Вышэйшая школа, 2012. — 720 с. — Режим доступа: <http://www.iprbookshop.ru/20289.html>

15 Сетубал, Ж. Введение в вычислительную молекулярную биологию [Текст]: пер. с англ. / Ж. Сетубал, Ж. Мейданис. - М.; Ижевск: Ин-т компьютер.исслед.: Dynamics, 2007. - 420 с. — Режим доступа: библиотечный фонд ЛИИ СО РАН.

16 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. — Режим доступа: <http://www.iprbookshop.ru/47719.html>

17 Степанова, С.М. Статистика [Электронный ресурс]: учебник / С.М. Степанова, Н.А. Рухманова, Т.Ю. Сорокина. — Электрон.текстовые данные. — СПб.: Интермедия, 2017. — 391 с. — Режим доступа: <http://www.iprbookshop.ru/66800.html>

18 Улитина, Е.В. Статистика [Электронный ресурс]: учебное пособие / Е.В. Улитина, О.В. Леднева, О.Л. Жирнова. — Электрон.текстовые данные. — М.: Московский финансово-промышленный университет «Синергия», 2013. — 320 с. — Режим доступа: <http://www.iprbookshop.ru/17045.html>

19 **Уэс, М.** Python и анализ данных [Электронный ресурс] / М. Уэс. — Электрон.текстовые данные. — Саратов: Профобразование, 2017. — 482 с. — Режим доступа: <http://www.iprbookshop.ru/64058.html>

20 **Двойнишников, С. В.** Методы обработки данных в научных исследованиях : учебное пособие / С. В. Двойнишников. — Новосибирск : Новосибирский государственный университет, 2022. — 76 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — Режим доступа: <https://www.iprbookshop.ru/128134.html>

7.2 Программное обеспечение

1. MicrosoftOffice
2. OpenOffice
3. MicrosoftWindows
4. AdobeAcrobatPro
5. Dr. WebCorporateAnti-Virus
6. KasperskyAnti-Virus
7. CorelDraw
8. GIMP
9. MrBayes
10. BEAST
11. BLAST
12. CLUSTAL
13. Программная среда R
14. Программная среда Python
15. Программная среда Java
16. Программная среда Perl
17. Inkscape
18. Gnuplot

7.3 Базы данных, информационно-справочные и поисковые системы

<http://www.biometrica.tomsk.ru/> - форум, обсуждающий различные вопросы применения статистических методов в научных исследованиях.

<https://cran.r-project.org/> - *хранилище пакетов для статического анализа с помощью языка программирования R*

<http://www.talkstats.com/> - международный форум по различным вопросам математической статистики

<https://www.freemathhelp.com/forum/> - международный форум по различным вопросам математики и математического моделирования

<http://www.bookre.org> – электронная библиотека рунета, поиск журналов и книг.

<http://elibrary.ru/defaultx.asp> – научная электронная библиотека, крупнейший российский информационный портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты более 12 млн научных статей и публикаций.

8. Материально-техническое обеспечение дисциплины (модуля):

Материально-техническое обеспечение института, необходимое для реализации программы включает в себя:

- Конференц-залы
- Мультимедийные установки, компьютерная техника с возможностью подключения к сети "Интернет".

9. Образовательные технологии:

При реализации различных видов учебной работы дисциплины используются следующие формы проведения занятий.

Стандартные методы обучения:

- Лекция;
- Видео-лекция;
- Дискуссия, круглый стол;
- Практические занятия;
- Самостоятельная работа;
- Консультации специалистов.

Обучения с применением интерактивных форм образовательных технологий:

- информационно-коммуникационные образовательные технологии - лекция-визуализация, представление научно-исследовательских работ с использованием специализированных программных сред;
- выполнение вычислений с использованием прикладных программ.

10 Кадровое обеспечение дисциплины (модуля)

Реализацию образовательного процесса по программе дисциплины обеспечивает старший научный сотрудник лаборатории геносистематики, кандидат биологических наук Букин Юрий Сергеевич.

Разработчик программы: к.б.н. Ю.С. Букин

11 Оценочные средства

Оценочные средства представлены в Приложении к рабочей программе дисциплины в виде фонда оценочных средств для проведения текущего контроля успеваемости и промежуточной аттестации аспирантов по освоению дисциплины.

ПРИЛОЖЕНИЕ А

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине (модулю)

«Методы статистической обработки данных»

ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины.

1 Оценочные средства текущего контроля

Текущий контроль проводится для оценки степени усвоения аспирантами учебных материалов, обозначенных в рабочей программе, и контроля СРС. Текущий контроль осуществляется в виде систематической проверки знаний и навыков аспирантов. Для этого используется устный опрос.

Контрольные вопросы для текущей аттестации:

Раскрыть следующие темы:

1. Численные методы математического анализа.
2. Численное решение линейных и нелинейных уравнений.
3. Численное решение дифференциальных уравнений.
4. Параметрические методы оценки параметров распределений.
5. Непараметрические методы оценки параметров распределений.
6. Параметрические методы оценки параметров распределений.
7. Проверка статистических гипотез с помощью стандартных статистических методов.
8. Методы дисперсионного анализа.
9. Методы корреляционного анализа.
10. Регрессионный анализ с помощью метода наименьших квадратов
11. Кластерный анализ
12. Многомерное шкалирование.
13. Анализ методом главных компонент.
14. Байесовское оценивание параметров распределений.
15. Байесовское тестирование статистических гипотез.

Критерии оценивания:

При оценке ответа учитывается:

- 1) полнота и правильность ответа;

- 2) степень осознанности, понимания изученного;
- 3) языковое оформление ответа.

Ответ оценивается на **«отлично»**, если аспирант: полно излагает изученный материал, дает правильное определенное понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только из литературы, но и самостоятельно составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка.

Ответ оценивается на **«хорошо»**, если аспирант даёт ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1-2 ошибки, которые сам же исправляет, и 1-2 недочёта в последовательности и языковом оформлении излагаемого.

«Удовлетворительно» ставится, если аспирант обнаруживает знание и понимание основных положений темы, но при этом: излагает материал неполно и допускает неточности в определении понятий или формулировке теорий; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.

Оценка **«неудовлетворительно»** ставится, если ответ не удовлетворяет требованиям положительной оценки или аспирант отказывается отвечать на контрольные вопросы.

2 Оценочные средства для промежуточной аттестации

Промежуточная аттестация проходит в форме зачета.

Контрольные вопросы для зачета:

1. Численное решение систем линейных уравнений.
2. Численное решение систем не линейных линейных уравнений.
3. Численные методы исследования функций.
4. Численные методы интегрирования дифференциальных уравнений.
5. Стационарные решения для дифференциальных уравнений первого порядка.
6. Стационарные решения для систем дифференциальных уравнений первого порядка.
7. Понятие вероятности и случайной величины.
8. Законы распределения случайных величин.
9. Генеральная совокупность и выборочные совокупности.
10. Оценка параметров распределений случайных величин.
11. Бутстреп методы для оценки параметров распределений.
12. Использование критерия Стьюдента для сравнения средних значений выборок.
13. Использование U-критерия Уилкоксона-Манна-Уитни для сравнения средних значений выборок.
14. Использование критерия Колмогорова-Смирнова для сравнения функций распределения

15. Критерий хи квадрат и сравнения распределений качественных признаков
16. Бутстреп метод для тестирования статистических гипотез
17. Дисперсионный анализ
18. Корреляционный анализ
19. Регрессионный анализ методом наименьших квадратов
20. Систематизация многомерных статистических данных
21. Кластерный анализ.
22. Многомерное шкалирование.
23. Анализ методом главных компонент.
24. Метод максимального правдоподобия для оценки параметров распределений.
25. Метод максимального правдоподобия для тестирования статистических гипотез.
26. Байесовский статистический анализ для оценки параметров распределений.
27. Теория планирования экспериментов и статистический анализ
28. Математические методы при исследовании динамики популяций
29. Вычислительная экология
30. Методы многомерной статистики в вычислительной географии, химии.
31. Тестирование статистических гипотез.
32. Методы теории случайных процессов.

Критерии оценки:

Оценивание аспиранта на промежуточной аттестации в форме зачета

Оценка зачета	Требования к знаниям и критерии выставления оценок
<i>Зачтено</i>	Аспирант при ответе демонстрирует большую часть содержания тем учебной дисциплины, владеет основными понятиями.
<i>Не зачтено</i>	Аспирант при ответе демонстрирует знание меньшей части содержания тем учебной дисциплины

ЛИСТ ОБНОВЛЕНИЯ

Дата	Внесенные обновления	Подпись