

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования «Новосибирский национальный исследовательский
государственный университет» (Новосибирский государственный университет, НГУ)

Факультет естественных наук

Согласовано

Декан ФЕН

Резников В.А.

подпись

«___» _____ 20__ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

МНОГОМЕРНЫЙ АНАЛИЗ БИОЛОГИЧЕСКИХ ДАННЫХ

направление подготовки: 06.04.01 Биология

направленность (профиль): Информационная биология

Форма обучения: очная

Разработчики:

д.б.н., профессор Ефимов В.М.

Зав. каф. информационной биологии
академик РАН Колчанов Н.А.

Руководитель программы:

д.б. н., профессор Шестопалова Л.В.

Новосибирск, 2020

Содержание

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы	3
2. Место дисциплины в структуре образовательной программы	5
3. Трудоемкость дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающегося с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающегося	5
4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий.....	6
5. Перечень учебной литературы	7
6. Перечень учебно-методических материалов по самостоятельной работе обучающихся..	7
7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	7
8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине	8
9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине	8
10. Оценочные средства для проведения текущего контроля и промежуточной аттестации по дисциплине.....	8
Приложение 1 Аннотация по дисциплине	
Приложение 2 Оценочные средства по дисциплине	

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Результаты освоения образовательной программы (компетенции)	В результате изучения дисциплины обучающиеся должны:		
	знать	уметь	владеть
ОПК-3 Готовность использовать фундаментальные биологические представления в сфере профессиональной деятельности для постановки и решения новых задач		- использовать стандартные пакеты статистических программ при построении интегральных показателей и отборе наиболее информативных переменных, и снижении размерностей анализируемых моделей	
ОПК-4 Способность самостоятельно анализировать имеющуюся информацию, выявлять фундаментальные проблемы, ставить задачу и выполнять полевые, лабораторные биологические исследования при решении конкретных задач с использованием современной аппаратуры и вычислительных средств, нести ответственность за качество работ и научную достоверность результатов		- применять как линейные, так и нелинейные методы анализа взаимного расположения объектов в многомерном пространстве и интерпретировать получаемые результаты с биологической точки зрения	
ОПК-5 Способность применять знание истории и методологии	- основные методы многомерного анализа данных: метод главных		

Результаты освоения образовательной программы (компетенции)	В результате изучения дисциплины обучающиеся должны:		
	знать	уметь	владеть
биологических наук для решения фундаментальных профессиональных задач	компонент, факторный анализ, дискриминантный анализ, регрессионные методы, многомерное шкалирование		
ПК-1 способность творчески использовать в научной и производственно-технологической деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность (профиль) программы магистратуры			–основными методами анализа одномерных и многомерных данных: вычисление квантилей, средних, среднеквадратичных отклонений, дисперсий, коэффициентов корреляции, построений графиков, ранжирований, центрирований и нормирований признаков, вычислений коэффициентов сходства/различия (расстояний) между объектами, кластерного анализа, метода главных компонент, факторного анализа, дискриминантного анализа, регрессионными методами, многомерным шкалированием, PLS-методами
ПК-3 способность применять методические основы		- применять как линейные, так и нелинейные методы анализа взаимного	

Результаты освоения образовательной программы (компетенции)	В результате изучения дисциплины обучающиеся должны:		
	знать	уметь	владеть
проектирования, выполнения полевых и лабораторных биологических, экологических исследований, использовать современную аппаратуру и вычислительные комплексы (в соответствии с направленностью (профилем) программы магистратуры)		расположения объектов в многомерном пространстве и интерпретировать получаемые результаты с биологической точки зрения	

2. Место дисциплины в структуре образовательной программы

Дисциплины (практики), изучение которых необходимо для освоения дисциплины
Многомерный анализ биологических данных:

- Биометрия
- Математический анализ
- Математическая статистика
- Высшая алгебра

Дисциплины (практики), для изучения которых необходимо для освоения дисциплины
Многомерный анализ биологических данных:

- Экспериментальные и компьютерные методы анализа биоразнообразия;
- Анализ биологических изображений;
- Математические основы системной биологии: моделирование молекулярно-генетических систем

а также при подготовке дипломной работы.

3. Трудоемкость дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающегося с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающегося

Трудоемкость дисциплины – 3 з.е. (108 ч)

Форма промежуточной аттестации: 2 семестр – экзамен

№	Вид деятельности	Семестр
		2
1	Лекции, ч	16
2	Практические занятия, ч	20

3	Лабораторные занятия, ч	-
4	Занятия в контактной форме, ч , из них	40
5	из них аудиторных занятий, ч	36
6	в электронной форме, ч	-
7	консультаций, час.	2
8	промежуточная аттестация, ч	2
9	Самостоятельная работа, час.	68
10	Всего, ч	108

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

2 семестр
Лекции (16 ч)

Наименование темы и их содержание	Объем, час
Раздел 1 Введение	
1. Необходимость многомерной обработки биологических данных. Геометрический подход: анализ расположения объектов в многомерном пространстве и направлений их изменчивости через корреляции с признаками. История (Ф.Гальтон, К.Пирсон, Р.Фишер, Г.Хотеллинг). Современное состояние: главные компоненты (факторный анализ), множественная регрессия, дискриминантный анализ, канонический анализ, многомерное шкалирование, PLS-методы. Возможность визуализации. Оценка достоверности и ее роль.	2
Раздел 2 Предварительная работа с данными в популяционных исследованиях	
1. Объекты. Признаки - свойства объектов, позволяющие отличать их друг от друга и измерять расстояние между ними. Типы признаков. Допустимые преобразования и сравнения. Средние и дисперсии выборки. Нормировки.	2
Раздел 3 Линейная алгебра	
1. Скаляры, вектора. матрицы. Евклидово пространство, точки, вектора, наборы векторов. Евклидово расстояние между точками, углы между векторами. Операции сложения и умножения. ортогональные, диагональные и единичные матрицы.	1
2. Преобразования: перенос, поворот, растяжение. Центроиды, дисперсия. Корреляционная матрица. Собственные вектора. Главные компоненты. Повороты (факторный анализ)	1
Раздел 4 Внутривыборочная изменчивость.	
1. Многомерный анализ как средство поиска биологического смысла при анализе изменчивости биологических объектов. Методы исследования: главные компоненты, факторный анализ. Отсечение дальних компонент. Примеры.	2
Раздел 5 Межвыборочная изменчивость.	
1. t-критерий. Дискриминантный анализ. Проблема коллинеарности. Метод Царапкина. Объединенная внутривыборочная изменчивость. Предварительная обработка методом главных компонент. Дискриминантный PLS-анализ.	2
Раздел 6 Внешние факторы как возможные причины изменчивости.	

1. Линейная регрессия. Проекция. Проблема коллинеарности. Регрессия на главные компоненты. PLS- регрессия.	2
Раздел 7 Нелинейные методы.	
1. Неевклидовы расстояния и меры сходства-различия. Многомерное шкалирование.	2
Раздел 8 Временные ряды.	
1. Теорема Такенса. Фазовые портреты. Гладкие и главные компоненты временных рядов. Методы прогноза временных рядов. Примеры.	2

Практические занятия (20 ч)

Содержание практического занятия	Объем, час
Семинары по темам лекционных занятий	20

Самостоятельная работа студентов (68 ч)

Перечень занятий на СРС	Объем, час
Подготовка к практическим занятиям.	56
Подготовка к экзамену	12

5. Перечень учебной литературы

5.1 Основная литература

1. Рокицкий П.Ф. Биологическая статистика. Минск: Высшая школа. 1973.
2. Глотов Н.В. и др. Биометрия. Ленинград: ЛГУ. 1982.
3. Васильева Л.А. Биологическая статистика. Новосибирск: ИЦиГ СО РАН. 2004.
4. Васильева Л.А. Статистические методы в биологии, медицине и сельском хозяйстве. Новосибирск: ИЦиГ СО РАН. 2007.

5.2 Дополнительная литература

5. Шеффе Н. Дисперсионный анализ. М: Физматгиз 1963.
6. Урбах В.Ю. Биометрические методы. М: Физматгиз.1964.
7. Ван дер Варден. Математическая статистика. М: ИЛ. 1960.
8. Снедекор Дж.У. Статистические методы в приложении к исследованиям в сельском хозяйстве и биологии. М: Сельхозиздат. 1961.

6. Перечень учебно-методических материалов по самостоятельной работе обучающихся

9. Ефимов В.М., Ковалева В.Ю. Многомерный анализ биологических данных: учебное пособие. 2-е испр. и доп. изд. – Санкт-Петербург: ВИЗР РАСХН, 2008. –87с.
10. Презентации лекций на сайте КИБ http://kib.nsu.ru/?page_id=564

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

Освоение дисциплины используются следующие ресурсы:

- электронная информационно-образовательная среда НГУ (ЭИОС);
- образовательные интернет-порталы;
- информационно-телекоммуникационная сеть Интернет.

Взаимодействие обучающегося с преподавателем (синхронное и (или) асинхронное) осуществляется через личный кабинет студента в ЭИОС, электронную почту.

7.1 Современные профессиональные базы данных:

Не используются

7.2 Информационные справочные системы

Не используются

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

8.1 Перечень программного обеспечения

ОС Windows и MicrosoftOffice

8.2 Информационные справочные системы

Не используются

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Для реализации дисциплины Многомерный анализ биологических данных используются специальные помещения:

1. Учебные аудитории для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля, промежуточной и итоговой аттестации;

2. Помещения для самостоятельной работы обучающихся;

Учебные аудитории укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду НГУ.

Для проведения занятий лекционного типа предлагаются следующие наборы демонстрационного оборудования и учебно-наглядных пособий:

- комплект лекций-презентаций по темам дисциплины.

Материально-техническое обеспечение образовательного процесса по дисциплине для обучающихся из числа лиц с ограниченными возможностями здоровья осуществляется согласно «Порядку организации и осуществления образовательной деятельности по образовательным программам для инвалидов и лиц с ограниченными возможностями здоровья в Новосибирском государственном университете».

10. Оценочные средства для проведения текущего контроля и промежуточной аттестации по дисциплине

Перечень результатов обучения по дисциплине Многомерный анализ биологических данных и индикаторов их достижения представлен в виде знаний, умений и владений в разделе 1.

10.1 Порядок проведения текущего контроля и промежуточной аттестации по дисциплине

Текущий контроль успеваемости:

Текущая аттестация по дисциплине «Многомерный анализ биологических данных» предусматривает контроль посещаемости занятий.

Промежуточная аттестация:

Промежуточная аттестация (итоговая по дисциплине) проводится по завершению периода ее освоения (семестра). Промежуточная аттестация по дисциплине включает 2 этапа:

- 1) сдача отчетов о выполнении практических заданий,
- 2) экзамен.

Для того чтобы быть допущенным к экзамену, студент должен выполнить следующее:

- в ходе прохождения дисциплины посетить не менее 60 % занятий;
- сдать все отчеты о выполнении практических заданий.

Выполнение практических заданий в полном соответствии с предъявляемыми требованиями к ее содержанию и срокам защиты (оценка «зачтено») является необходимым условием допуска к экзамену.

Описание критериев и шкал оценивания индикаторов достижения результатов обучения по дисциплине Многомерный анализ биологических данных

Таблица 10.1

Код компетенции	Результат обучения по дисциплине	Оценочное средство
ОПК-3	Умение использовать стандартные пакеты статистических программ при построении интегральных показателей и отборе наиболее информативных переменных, и снижении размерностей анализируемых моделей	Выполнение практических заданий Экзамен
ОПК-4	Умение применять как линейные, так и нелинейные методы анализа взаимного расположения объектов в многомерном пространстве и интерпретировать получаемые результаты с биологической точки зрения.	Выполнение практических заданий Экзамен
ОПК-5	Знание основных методов многомерного анализа данных: метода главных компонент, факторного анализа, дискриминантного анализа, регрессионных методов, многомерного шкалирования	Экзамен

ПК-1	Владение основными методами анализа одномерных и многомерных данных: вычисление квантилей, средних, среднеквадратичных отклонений, дисперсий, коэффициентов корреляции, построений графиков, ранжирований, центрирований и нормирований признаков, вычислений коэффициентов сходства/различия (расстояний) между объектами, кластерного анализа, метода главных компонент, факторного анализа, дискриминантного анализа, регрессионными методами, многомерным шкалированием, PLS-методами	Выполнение практических заданий Экзамен
ПК-3	Умение применять как линейные, так и нелинейные методы анализа взаимного расположения объектов в многомерном пространстве и интерпретировать получаемые результаты с биологической точки зрения	Выполнение практических заданий Экзамен

Таблица 10.2

Критерии оценивания результатов обучения	Шкала оценивания
Экзамен: – свободное владение теоретическим и фактическим материалом, – полнота понимания используемых для обработки результата методов, – самостоятельность, осмысленность, структурированность, логичность и аргументированность изложения материала, отсутствие затруднений при формулировке собственных суждений, – точность и корректность применения терминов и понятий, – наличие исчерпывающих ответов на дополнительные вопросы. При изложении ответа на вопрос(ы) экзаменационного билета обучающийся мог допустить не принципиальные неточности.	<i>Отлично</i>
Экзамен: – владение теоретическим и фактическим материалом, – полнота понимания используемых для обработки результата методов, – самостоятельность, осмысленность, структурированность, логичность и аргументированность изложения материала, наличие затруднений при формулировке собственных суждений, – точность и корректность применения терминов и понятий при наличии незначительных ошибок, – наличие полных ответов на дополнительные вопросы с возможным присутствием ошибок.	<i>Хорошо</i>
Экзамен: – слабое владение теоретическим и фактическим материалом, – частичное понимание используемых для анализа методов, – самостоятельность и осмысленность в изложении материала, наличие ошибок в логике и аргументации, а также затруднений при формулировке собственных суждений,	<i>Удовлетворительно</i>

– корректность применения терминов и понятий при наличии незначительных ошибок, – наличие неполных и/или содержащих существенные ошибки ответов на дополнительные вопросы.	
Экзамен: – фрагментарное и недостаточное владение теоретическим и фактическим материалом, – отсутствие осмысленности, структурированности, логичности и аргументированности в изложении материала, – грубые ошибки в применении терминов и понятий, – отсутствие ответов на дополнительные вопросы.	<i>Неудовлетворительно</i>
Выполнение практических заданий – не менее 50% ответов должны быть правильными.	<i>зачтено</i>
Выполнение практических заданий – присутствие многочисленных ошибок (более 70% ответов содержат ошибки).	<i>незачтено</i>

Типовые контрольные задания и иные материалы, необходимые для оценки результатов обучения

Оценочные материалы по промежуточной аттестации (приложение 2), предназначенные для проверки соответствия уровня подготовки по дисциплине требованиям ФГОС, хранятся на кафедре-разработчике РПД в печатном и электронном виде.

Темы для текущего контроля:

1. Пакет PAST. Построение графиков. Вычисление статистических параметров. Вычисление достоверности разности средних. Стандартизация признаков.
2. Пакет PAST. Корреляция признаков. Главные компоненты. Евклидовы расстояния. Кластерный анализ. Неметрическое шкалирование.
3. Пакет R. Построение графиков. Вычисление статистических параметров. Стандартизация признаков. Корреляция признаков. Главные компоненты. Пакет JACOBI-4. Евклидовы расстояния. Кластерный анализ. Неметрическое шкалирование.

Вопросы для экзамена:

- Вопрос 1. Необходимость многомерной обработки биологических данных
- Вопрос 2. Геометрический подход: анализ расположения объектов в многомерном пространстве и направлений их изменчивости через корреляции с признаками.
- Вопрос 3. История и современное состояние многомерной обработки биологических данных.
- Вопрос 4. Основные методы анализа биологических данных: главные компоненты (факторный анализ), множественная регрессия, дискриминантный анализ, канонический анализ, многомерное шкалирование, PLS-методы.
- Вопрос 5. Многомерный анализ как средство поиска биологического смысла при анализе изменчивости биологических объектов.
- Вопрос 6. Объекты. Признаки - свойства объектов, позволяющие отличать их друг от друга и измерять расстояние между ними. Типы признаков.
- Вопрос 7. Скаляры, вектора. матрицы.

Вопрос 8. Методы работы с матрицами.

Вопрос 9. Метод главных компонент.

Вопрос 10. t-критерий. Дискриминантный анализ.

Вопрос 11. PLS- регрессия.

Вопрос 12. Неевклидовы расстояния и меры сходства-различия. Многомерное шкалирование.

Вопрос 13. Теорема Такенса..

Вопрос 14. Фазовые портреты.

Вопрос 15. Гладкие и главные компоненты временных рядов. Методы прогноза временных рядов.

Вопрос 16. Проблема коллинеарности. Регрессия на главные компоненты

**Лист актуализации рабочей программы дисциплины
«Многомерный анализ биологических данных»**

№	Характеристика внесенных изменений (с указанием пунктов документа)	Дата и № протокола Ученого совета ФЕН	Подпись ответственного