

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное
образовательное учреждение высшего образования_
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Институт биологии и биомедицины

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Рабочая программа дисциплины

Анализ данных в биологических дисциплинах

Уровень высшего образования

Магистратура

Направление подготовки / специальность

06.04.01 - Биология

Направленность образовательной программы

Молекулярная биология и иммунология

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.О.02 Анализ данных в биологических дисциплинах относится к обязательной части образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ОПК-6: Способен творчески применять и модифицировать современные компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок;	ОПК-6.1: Знает: - пути и перспективы применения современных компьютерных технологий в биологических науках и образовании ОПК-6.2: Умеет: - работать с профессиональными базами и банками данных в избранной области профессиональной деятельности ОПК-6.3: Владеет: - необходимым математическим аппаратом и навыками анализа и хранения электронных изображений, имеет опыт модификации компьютерных технологий в целях профессиональных исследований	ОПК-6.1: Знает современные компьютерные технологии и программы для анализа данных, правила работы и синтаксис языка для статистических вычислений R ОПК-6.2: Уметь работать с разными типами данных с использованием языка R, загружать и манипулировать таблицами данных, выбирать подходящий способ визуализации данных ОПК-6.3: Владеет навыками проведения анализа данных с использованием языка R, построения статистических моделей и использования статистических тестов на языке R, интерпретации и визуализации результатов исследования с помощью графической подсистемы языка R	Расчетно-графическое задание	Зачёт: Расчетно-графическое задание
ОПК-7: Способен в сфере своей профессиональной деятельности самостоятельно определять	ОПК-7.1: Знает: - основные источники и методы получения профессиональной информации, направления научных исследований,	ОПК-7.1: Знает принципы планирования эксперимента; основные математические законы, лежащие в основе актуальных	Тест	Зачёт: Тест

стратегию и проблематику исследований, принимать решения, в том числе инновационные, выбирать и модифицировать методы, отвечать за качество работ и внедрение их результатов, обеспечивать меры производственной безопасности при решении конкретной задачи;	соответствующих направленности программы магистратуры ОПК-7.2: Умеет: - выявлять перспективные проблемы и формулировать принципы решения актуальных научно-исследовательских задач на основе использования комплексной информации, в том числе на стыке областей знания; - разрабатывать методики решения и координировать выполнение отдельных заданий при руководстве группой исследователей, с учетом требований техники безопасности ОПК-7.3: Владеет: - методами анализа достоверности и оценки перспективности результатов проведенных экспериментов и наблюдений; - опытом обобщения и анализа научной и научно-технической информации; - опытом представления полученных результатов в виде докладов и публикаций	подходов к анализу данных; правила выбора статистических критериев для тестирования гипотез; принципы построения математических моделей и прогнозов в биологических исследованиях ОПК-7.2: Умеет выбирать соответствующие типу данных критерии и методы математической статистики, интерпретировать результаты анализа ОПК-7.3: Владеет навыками создания математических моделей, навыками построения прогнозов на основе созданных моделей		
--	---	---	--	--

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	2
Часов по учебному плану	72
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	14
- занятия семинарского типа (практические занятия / лабораторные работы)	14
- КСР	1
самостоятельная работа	43
Промежуточная аттестация	0 Зачёт

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабора- торные работы), часы	Всего	
	о ф о	о ф о	о ф о	о ф о	о ф о
Тема 1. Фундаментальные основы математической статистики	14	2	2	4	10
Тема 2. Общие линейные модели	20	4	4	8	12
Тема 3. Обобщенные линейные модели	10	2	2	4	6
Тема 4. Общие аддитивные модели	10	2	2	4	6
Тема 5. Элементы многомерного анализа	9	2	2	4	5
Тема 6. Бутстреп и рандомизация	8	2	2	4	4
Аттестация	0				
КСР	1				1
Итого	72	14	14	29	43

Содержание разделов и тем дисциплины

Тема 1. Фундаментальные основы математической статистики

Тема 2. Общие линейные модели

Тема 3. Обобщенные линейные модели

Тема 4. Общие аддитивные модели

Тема 5. Элементы многомерного анализа

Тема 6. Бутстреп и рандомизация

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Самостоятельная работа студентов направлена на самостоятельное изучение практических подходов к анализу и обработке данных биологических исследований.

Цель самостоятельной работы – подготовка современного компетентного специалиста и формирование способностей и навыков к непрерывному самообразованию и профессиональному совершенствованию.

Виды самостоятельной работы студентов в рамках освоения дисциплины:

- изучение понятийного аппарата и проработка тем дисциплины;
- работа с основной и дополнительной литературой дома и в библиотеке;
- изучение сайтов по темам дисциплины в сети Интернет;

- подготовка к тестам;
- подготовка к решению задач;
- подготовка к зачету.

Изучение понятийного аппарата дисциплины

Вся система индивидуальной самостоятельной работы должна быть подчинена усвоению понятийного аппарата, поскольку одной из важнейших задач подготовки современного грамотного специалиста является овладение и грамотное применение профессиональной терминологии. Лучшему усвоению и пониманию дисциплины помогут различные энциклопедии, словари, справочники и другие материалы, указанные в списке литературы.

Работа над основной и дополнительной литературой

Изучение рекомендованной литературы следует начинать с учебников и учебных пособий, затем переходить к нормативно-правовым актам, научным монографиям и материалам периодических изданий. Конспектирование – одна из основных форм самостоятельного труда, требующая от студента активно работать с учебной литературой и не ограничиваться конспектом лекций.

Студент должен уметь самостоятельно подбирать необходимую для учебной и научной работы литературу. При этом следует обращаться к предметным каталогам и библиографическим справочникам, которые имеются в библиотеках.

Для аккумуляции информации по изучаемым темам рекомендуется формировать личный архив, а также каталог используемых источников. При этом если уже на первых курсах обучения студент определяет для себя наиболее интересные сферы для изучения, то подобная работа будет весьма продуктивной с точки зрения формирования библиографии для последующего написания дипломного проекта на выпускном курсе.

Изучение сайтов по темам дисциплины в сети Интернет

Ресурсы Интернет являются одним из альтернативных источников быстрого поиска требуемой информации. Их использование возможно для получения основных и дополнительных сведений по изучаемым материалам.

Самостоятельная работа студента при подготовке к зачету

Контроль выступает формой обратной связи и предусматривает оценку успеваемости студентов и разработку мер по дальнейшему повышению качества подготовки современных специалистов.

В начале семестра рекомендуется внимательно изучить перечень вопросов к зачету по данной дисциплине (представлен в разделе 6.4), а также использовать в процессе обучения программу, другие методические материалы, разработанные кафедрой по данной дисциплине. Это позволит в процессе изучения тем сформировать более правильное и обобщенное видение студентом существа того или иного вопроса за счет:

- а) уточняющих вопросов преподавателю;
- б) самостоятельного уточнения вопросов на смежных дисциплинах;
- в) углубленного изучения вопросов темы по учебным пособиям.

После изучения соответствующей тематики рекомендуется проверить наличие и формулировки вопроса по этой теме в перечне вопросов к зачету, а также попытаться изложить ответ на этот вопрос. Если возникают сложности при раскрытии материала, следует вновь обратиться к

лекционному материалу, материалам практических занятий, уточнить терминологический аппарат темы, а также проконсультироваться с преподавателем.

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Расчетно-графическое задание) для оценки сформированности компетенции ОПК-6:

На 5 пробных площадях с разным типом почвы (А-Е) выращивали растения. В конце эксперимента измерили высоту всех выращенных растений. Полученные данные приведены в файле hw2.surname.data.xlsx.

1. Загрузите данные в R любым способом (при использовании вспомогательных файлов, приложите их к решению). Данные необходимо сформировать в виде фрейма из двух переменных - высота растений (количественная переменная), тип почвы (группирующая текстовая переменная либо фактор). Имена фрейма и переменных должны быть осмысленными.
2. Рассчитайте средние значения, стандартные отклонения и доверительные интервалы для средней высоты растений на площадках с разным типом почвы.
3. Постройте параллельные диаграммы размахов высоты растений для разных типов почвы. Отдельные диаграммы должны быть залиты, причем диаграмма для почвы с максимальной средней высотой растений должна быть красной, а для почвы с минимальной средней высотой - желтой.
4. Сформируйте модель дисперсионного анализа зависимости высоты растений от типа почвы.
5. Проанализируйте характер распределения остатков. Постройте гистограмму их распределения, примените три критерия согласия. В комментарии приведите р-значения и сделайте итоговый вывод.
6. Проведите анализ гомогенности дисперсии высоты растений на площадках с разным типом почвы. В комментарии приведите название процедуры, р-значение и сделайте вывод.
7. Выберите наиболее подходящую процедуру для анализа зависимости высоты растений от типа почвы. Проведите анализ. В комментарии приведите название процедуры, р-значение и сделайте вывод.
8. В случае выявления значимых отличий между высотой растений, выращенных на почве разного типа, проведите попарные сравнения для выявления конкретных типов почвы, отличия между которыми статистически значимы. Используйте процедуру с применением поправки Холма. В комментарии укажите пары типов почвы, отличия между которыми статистически значимы.
9. Проведите попарные сравнения с применением пошаговой процедуры Вестфолла. В комментарии укажите, для каких дополнительных пар типов почвы выявлены статистически значимые отличия по сравнению с процедурой на основе поправки Холма.

Критерии оценивания (оценочное средство - Расчетно-графическое задание)

Оценка	Критерии оценивания
зачтено	50-100% правильно выполненных задач

Оценка	Критерии оценивания
не зачтено	менее 50% правильно выполненных задач

5.1.2 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ОПК-7:

1. Кто разработал метод дисперсионного анализа?

- а) Питер Стьюдент;
- б) Уильям Госсет;
- в) Рональд Фишер;
- г) Карл Пирсон.

2. Дисперсионный анализ основан на:

- а) отношении среднего к дисперсии;
- б) произведении чисел степеней свободы;
- в) отношении дисперсии к среднему;
- г) отношении средних квадратов.

3. При сравнении 4 групп по 10 наблюдений числа степеней свободы в дисперсионном анализе составят:

- а) 3 и 9;
- б) 4 и 18;
- в) 5 и 18;
- г) 3 и 36.

4. Чем больше F-отношение в дисперсионном анализе:

- а) тем больше число степеней свободы;
- б) тем выше гомогенность дисперсии;
- в) тем вероятнее наличие отличий между средними;
- г) тем больше внутригрупповая дисперсия.

5. В стандартной таблице дисперсионного анализа не указывается:

- а) источник изменчивости;
- б) средние квадраты;
- в) F-отношение;
- г) объем выборки.

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	50-100% правильных ответов
не зачтено	менее 50% правильных ответов

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено			зачтено			
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые	Продemonстрированы все основные умения. Решены все основные задачи с отдельными несущественными	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов

			полном объеме	объеме, но некоторые с недочетами	с недочетами .	недочетам и, выполнен ы все задания в полном объеме	
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрирован творческий подход к решению нестандартных задач

Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Расчетно-графическое задание) для оценки сформированности компетенции ОПК-6

Описание данных. Набор данных iris содержит данные о длине и ширине чашелистиков (переменные Sepal.Length и Sepal.Width), а также о длине и ширине лепестков (переменные Petal.Length и Petal.Width) в цветках трех видов ирисов: *Iris setosa*, *I. versicolor* и *I. virginica* (переменная Species: принимает значения setosa, versicolor и virginica соответственно).

1. Загрузите данные по ирисам. Указание: воспользуйтесь функцией data().
2. Проведите статистический анализ ширины чашелистика у разных видов ириса. Рассчитайте средние и стандартные отклонения ширины чашелистика для всех видов. Есть ли отличия между видами? Между какими парами видов есть статистически значимые отличия и в какую сторону? Обоснуйте выбор метода анализа, при использовании статистических критериев в комментарии приведите р-значения и сформулируйте вывод.
3. Рассчитайте 99.9 %-ный доверительный интервал медианы ширины чашелистика для каждой породы ирисов с использованием бутстрепа. Используйте метод с коррекцией смещения и скошенности.
4. Постройте итоговый график: столбчатая диаграмма, отображающая медианы ширины чашелистика у разных видов ириса с доверительными интервалами. Виды ирисов должны следовать в порядке убывания ширины чашелистика. Оси должны быть подписаны, столбики для разных видов ирисов должны отличаться цветом. Разместите на графике компактные буквенные обозначения, отражающие значимость отличий между видами ирисов.

Критерии оценивания (оценочное средство - Расчетно-графическое задание)

Оценка	Критерии оценивания
зачтено	50-100% правильно выполненных задач
не зачтено	менее 50% правильно выполненных задач

5.3.2 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ОПК-7

1. Чем нормальное распределение отличается от распределения Стьюдента?
 - а) оно предназначено для описания распределения дискретной случайной величины;
 - б) у него нет дополнительного параметра – числа степеней свободы;
 - в) оно характеризуется большим числом наблюдений в «хвостах»;
 - г) оно несимметрично.
2. Выберите пример несвязанных выборок:
 - а) значения диаметров деревьев в двух городских парках;
 - б) значения признака на правой и левой сторонах тела;
 - в) цены на учебники по биологии в двух онлайн-магазинах;
 - г) значения урожая пшеницы в регионах РФ до и после кризиса 2008 года.
3. При сравнении 5 групп по 20 наблюдений числа степеней свободы в дисперсионном анализе составят:
 - а) 5 и 50;
 - б) 4 и 95;**
 - в) 6 и 50;
 - г) 4 и 19.
4. Непараметрический аналог дисперсионного анализа:
 - а) критерий Крускала-Уоллиса;
 - б) критерий Манна-Уитни;
 - в) критерий Уилкоксона;
 - г) критерий Фридмана.

5. Сколько необходимо попарных сравнений при работе с 5 выборками?

- а) 4; б) 25; в) 5; г) 10.

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	50-100% правильных ответов
не зачтено	менее 50% правильных ответов

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Анализ данных : учебник / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. - Москва : Юрайт, 2023. - 490 с. - (Высшее образование). - ISBN 978-5-534-00616-2. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=847079&idb=0>.
2. Энатская Н. Ю. Математическая статистика и случайные процессы : учебное пособие / Н. Ю. Энатская. - Москва : Юрайт, 2023. - 201 с. - (Высшее образование). - ISBN 978-5-9916-9808-5. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=848905&idb=0>.

Дополнительная литература:

1. Гателюк О. В. Практикум по теории вероятностей и математической статистике : Учебное пособие для вузов / Гателюк О. В., Манюкова Н. В.; Манюкова Н. В. - Санкт-Петербург : Лань, 2022. - 132 с. - Книга из коллекции Лань - Математика. - ISBN 978-5-8114-9842-0., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=804174&idb=0>.
2. Горленко О. А. Дисперсионный анализ экспериментальных данных : учебное пособие / О. А. Горленко, Н. М. Борбаць, Т. П. Можеева. - 2-е изд. ; испр. и доп. - Москва : Юрайт, 2023. - 132 с. - (Высшее образование). - ISBN 978-5-534-14677-6. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=840028&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. <http://www.r-project.org/> – R Project: The R Project for Statistical Computing – язык программирования и среда разработки для статистических вычислений.
2. <http://cran.r-project.org/> – CRAN: The Comprehensive R Archive – архив пакетов для расширения языка программирования R.
3. <http://www.rstudio.com/> – R Studio – среда разработки для языка программирования R.
4. <http://www.statsoft.ru/home/textbook/> – электронный учебник по статистике и планированию эксперимента.
5. <http://r-analytics.blogspot.ru/> – R: Анализ и визуализация данных.

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами.

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

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 06.04.01 - Биология.

Автор(ы): Якимов Василий Николаевич, доктор биологических наук, доцент.

Заведующий кафедрой: Якимов Василий Николаевич, доктор биологических наук.

Программа одобрена на заседании методической комиссии от 05.12.2023г., протокол № 2.