

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Наумова Наталия Александровна
Должность: Ректор
Дата подписания: 08.07.2026 10:00:03
Уникальный программный код:
6b5279da4e034bffa79172803da5b71c0f0b0da

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ)

Факультет естественных наук
Кафедра общей биологии и биоэкологии

Согласовано
и.о. декана факультета естественных наук
« 14 » 04 2026 г.
/Лялина И.Ю./

Рабочая программа дисциплины

Основы математической статистики эколого-генетических систем на языке
программирования R

Направление подготовки
06.04.01 Биология

Программа подготовки:
Биоинформатика и анализ больших данных в экологическом мониторинге

Квалификация
Магистр

Форма обучения
Очная

Согласовано учебно-методической
комиссией факультета естественных наук
Протокол « 14 » 04 2026 г. № 6
Председатель УМКом /Лялина И.Ю./

Рекомендовано кафедрой общей биологии
и биоэкологии
Протокол от « 06 » 03 2026 г. № 8
Зав. кафедрой /Гордеев М.И./

Москва
2026

Авторы-составители:

Гордеев Михаил Иванович, доктор биологических наук, профессор
Москаев Антон Вячеславович, кандидат биологических наук, доцент
Темников Андрей Андреевич, старший преподаватель

Рабочая программа дисциплины «Основы математической статистики эколого-генетических систем на языке программирования R» составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 06.04.01 Биология, утвержденного приказом МИНОБРНАУКИ РОССИИ от 11.08.2020г., № 934

Дисциплина входит в часть, формируемую участниками образовательных отношений Блока 1 «Дисциплины (модули)» и является обязательной

Год начала подготовки (по учебному плану) 2026

Содержание

1.	Планируемые результаты обучения.....	4
1.1.	Цель и задачи дисциплины.....	4
1.2.	Задачи дисциплины:.....	4
1.3.	Планируемые результаты обучения	4
2.	Место дисциплины в структуре образовательной программы	5
3.	Объем и содержание дисциплины	6
3.1	объем дисциплины	6
3.2	содержание дисциплины	6
4.	Учебно-методическое обеспечение самостоятельной работы обучающихся	8
5.	Фонд оценочных средств для проведения текущей и промежуточной аттестации по дисциплине.....	11
5.1	перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.....	11
5.2	описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания	11
5.3	типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы.....	17
5.4	методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.	24
5.5	итоговая шкала выставления оценки по дисциплине	26
6.	Учебно-методическое и ресурсное обеспечение дисциплины.....	26
6.1	основная литература	26
6.2	дополнительная литература	26
6.3	ресурсы информационно-телекоммуникационной сети «интернет»	27
7.	Методические указания по освоению дисциплины	27
8.	Информационные технологии для осуществления образовательного процесса по дисциплине.....	27
9.	МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	30

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

1.1. Цель и задачи дисциплины

Цель освоения дисциплины: формирование у магистрантов системы знаний и практических навыков статистического анализа экологических и генетических данных с использованием языка R, необходимых для планирования эксперимента, проверки научных гипотез, построения и интерпретации современных статистических моделей в профессиональной исследовательской деятельности.

Задачи дисциплины:

Теоретико-статистический фундамент:

- Сформировать понимание логики статистического вывода, принципов проверки гипотез, понятий p уровня, ошибок I и II рода, мощности критерия и доверительных интервалов.
- Изучить теоретические основы классических параметрических и непараметрических методов (t тесты, ANOVA, χ^2), линейных и обобщённых линейных моделей (GLM).
- Освоить базовые концепции многомерного анализа (PCA, NMDS, иерархическая кластеризация) и популяционной генетики (закон Харди–Вайнберга, F статистики Райта).

Инструментальная подготовка в среде R:

- Приобрести устойчивые навыки работы в R и RStudio: импорт, очистка, преобразование и визуализация данных (пакеты tidyverse, ggplot2).
- Научить создавать полностью воспроизводимые аналитические отчёты средствами R Markdown, документируя весь ход анализа.

Методология анализа экологических данных:

- Освоить методы описательной статистики и сравнения групп с учётом структуры данных и проверки допущений.
- Научить строить, диагностировать и интерпретировать линейные регрессионные и ANOVA-модели, включая анализ взаимодействий факторов.
- Сформировать умение применять обобщённые линейные модели (логистическая, пуассоновская регрессия) к бинарным и счётным биологическим данным.
- Развить навыки многомерного анализа сообществ: ординация (PCA, NMDS, RDA), кластеризация, расчёт и интерпретация индексов альфа- и бета-разнообразия (пакет vegan).

Методология популяционно-генетического анализа:

- Научить вычислять и анализировать частоты аллелей и генотипов, тестировать отклонения от равновесия Харди–Вайнберга.
- Сформировать умение рассчитывать F статистики (F_{is} , F_{it} , F_{st}), оценивать генетическую дифференциацию и визуализировать популяционную структуру с помощью пакетов adegenet, hierfstat и poppr.

Научно-исследовательская компетенция:

- Развить способность самостоятельно формулировать статистическую гипотезу, обоснованно выбирать метод анализа, критически оценивать его ограничения и представлять результаты в форме, принятой в научных публикациях и отчётах.

1.2. Планируемые результаты обучения

В результате освоения данной дисциплины у обучающихся формируются следующие компетенции:

СПК-2. Способен использовать программы для эпигенетического анализа и качества эпигенетических данных;

СПК-4. Способен проводить экспертно-аналитическую работу при проведении научных исследований и экспериментальных работ

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина входит в часть, формируемую участниками образовательных отношений Блока 1 «Дисциплины (модули)» и является обязательной для изучения

Дисциплина «Основы математической статистики эколого-генетических систем на языке программирования R» опирается на знания, умения и виды деятельности, формируемые в процессе изучения дисциплин: «Биоинформатика (работа в базах данных)», «Биоинформационные подходы в геномике».

Дисциплина «Основы математической статистики эколого-генетических систем на языке программирования R» является методологической основой для исследовательской и теоретической работы в рамках подготовки магистерской диссертации.

3. ОБЪЕМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

3.1 Объем дисциплины

Показатель объема дисциплины	Форма обучения Очная
Объем дисциплины в зачетных единицах	2
Объем дисциплины в часах	72
Контактная работа:	14,2
Лекции	4
Практические занятия	10
Контактные часы на промежуточную аттестацию:	0,2
Зачёт	0,2
Самостоятельная работа	50
Контроль	7,8

Форма промежуточной аттестации: зачёт в 4 семестре

3.2 Содержание дисциплины

Наименование разделов (тем) Дисциплины с кратким содержанием	Количество часов	
	Лекции	Практические занятия
Тема 1. Статистический вывод и линейное моделирование в экологии Основное содержание: • Данные в экологии: типы переменных, распределения. Описательная статистика и визуализация (ggplot2). • Проверка статистических гипотез: параметрические (t-критерий) и непараметрические (Манна–Уитни) тесты. • Линейные модели: простая и множественная линейная регрессия, однофакторный и двухфакторный ANOVA, анализ взаимодействий. Диагностика моделей (остатки, выбросы, гетероскедастичность). • Обобщённые линейные модели (GLM): логистическая регрессия для бинарных данных, пуассоновская регрессия для счётных признаков. • Практические работы: описательная статистика и сравнение групп; регрессионный и дисперсионный анализ; построение GLM для экологических данных.	2	6
Тема 2. Многомерный анализ и элементы популяционной генетики Основное содержание: • Методы ординации сообществ: анализ главных компонент (PCA), неметрическое многомерное шкалирование (NMDS), избыточный анализ (RDA). Оценка и интерпретация индексов видового разнообразия (Шеннона, Симпсона). Пакет vegan в R. • Кластерный анализ: иерархическая кластеризация, метод k-средних. Визуализация тепловых карт и дендрограмм. • Основы популяционной генетики: частоты аллелей и генотипов, равновесие Харди–Вайнберга, тестирование отклонений. F-статистики Райта (Fis, Fit, Fst) как мера генетической дифференциации. Визуализация	2	4

генетической структуры популяций (пакеты adegenet, hierfstat). • Практические работы: ординация и кластеризация экологических сообществ; анализ генетической структуры популяций (частоты, HWE, F-статистики).		
Итого:	4	10

4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Темы для самостоятельного изучения	Изучаемые вопросы	Количество часов	Формы самостоятельной работы	Методическое обеспечение	Формы отчетности
Тема 1. Статистический вывод и линейное моделирование в экологии	- Современные методы выбора моделей: информационные критерии (AIC, BIC, AICc), кросс-валидация, регуляризация (LASSO, ridge regression) в R. - Смешанные линейные модели (LMM) и обобщённые смешанные модели (GLMM): учёт случайных эффектов и иерархической структуры данных в экологии (пакеты lme4, glmmTMB). - Анализ временных рядов в экологии: автокорреляция, модели ARIMA, сезонная декомпозиция, пакеты forecast, zoo. - Пространственная автокорреляция и её учёт в регрессионных моделях: индексы Морана, Geary, пакет spdep. - Байесовские методы в экологии: основы MCMC, пакеты brms, rstanarm,	25	Анализ литературных источников, конспектирование, подготовка к практическим занятиям	Основная и рекомендуемая учебная и научная литература, Интернет-ресурсы	Опрос, защита практикума

Темы для самостоятельного изучения	Изучаемые вопросы	Количество часов	Формы самостоятельной работы	Методическое обеспечение	Формы отчетности
	сравнение с частотными подходами. - Анализ выживаемости и времени до события в экологии: оценки Каплана–Мейера, регрессия Кокса, пакет survival. - Работа с отсутствующими данными: типы пропусков, методы импутации (mice, missForest), влияние на оценки параметров.				
Тема 2. Многомерный анализ и элементы популяционной генетики	- Канонический анализ соответствия (CCA) и детрендовый анализ соответствия (DCA): теория, сравнение с PCA и RDA, интерпретация ординационных диаграмм. - Анализ главных координат (PCoA) и неметрическое многомерное шкалирование (NMDS): выбор метрики расстояния (Bray–Curtis, Jaccard, Gower), оценка качества визуализации (stress). - PERMANOVA (adonis2) и	25	Анализ литературных источников, конспектирование, подготовка к практическим занятиям	Основная и рекомендуемая учебная и научная литература, Интернет-ресурсы	Опрос, защита практикума

Темы для самостоятельного изучения	Изучаемые вопросы	Количество часов	Формы самостоятельной работы	Методическое обеспечение	Формы отчетности
	анализ гомогенности многомерных дисперсий (betadisper) для сравнения групп сообществ. • Индексы бета-разнообразия и их разложение на вклад вложенности и замены видов (пакет betapart). • Анализ главных компонент по аллельным частотам (PCA для генетических данных), визуализация популяционной структуры с помощью DAPC (adegenet). • AMOVA (анализ молекулярной дисперсии) и популяционная дифференциация : расчёт Fst, Ф-статистики, оценка потока генов (пакеты poppr, hierfstat). • Геномные ландшафтные методы: выявление адаптивных локусов, тесты на селекцию (Fst-выбросы, BayeScan), связь генетической структуры с факторами среды.				

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

5.1 Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

Код и наименование компетенции	Этапы формирования
СПК-2 . Способен использовать программы для эпигенетического анализа и качества эпигенетических данных	1. Работа на учебных занятиях 2. Самостоятельная работа
СПК-4 . Способен проводить экспертно-аналитическую работу при проведении научных исследований и экспериментальных работ	1. Работа на учебных занятиях 2. Самостоятельная работа

5.2 Описание показателей и критериев оценивания компетенций на различных этапах формирования, описание шкал оценивания

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
СПК-2	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа	знать: – интерфейс и базовые возможности языка R и среды RStudio; – основные типы данных, структуры (векторы, матрицы, списки, data.frame) и приёмы их обработки; – форматы входных файлов (.csv, .tsv, .xlsx) и способы их импорта; – критерии качества исходных данных (пропуски, выбросы, несоответствие типов); – ключевые статистические функции и пакеты (базовый R, ggplot2). уметь: – загружать данные в R, выполнять первичный осмотр (summary, str, head); – выявлять и корректно обрабатывать пропуски и явные ошибки; – строить основные типы графиков (гистограммы, боксплоты, диаграммы рассеяния); – выполнять описательную статистику и базовые тесты (t-тест,	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			критерий χ^2).		
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа	знать: – интерфейс и базовые возможности языка R и среды RStudio; – основные типы данных, структуры (векторы, матрицы, списки, data.frame) и приёмы их обработки; – форматы входных файлов (.csv, .tsv, .xlsx) и способы их импорта; – критерии качества исходных данных (пропуски, выбросы, несоответствие типов); – ключевые статистические функции и пакеты (базовый R, ggplot2). уметь: – создавать воспроизводимые отчёты (R Markdown) с включением кода, результатов и интерпретации; – применять продвинутые графические системы (ggplot2 с настройкой тем, facets) для визуализации многомерных данных; – автоматизировать рутинные операции с помощью собственных функций и циклов; – критически оценивать качество исходных эпигенетических/экологических данных (проверка на артефакты, автокорреляцию) и выбирать подходящий статистический инструментарий. владеть: – навыками работы с пакетами для контроля качества и препроцессинга (dplyr, tidyr, janitor); – техникой составления аналитических скриптов, обеспечивающих полную воспроизводимость от	Защита практикума	Шкала оценивания защиты практикума

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			исходных файлов до финальных выводов.		
СПК-4	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа	знать: – логику проверки статистических гипотез, p-value и его интерпретацию; – основы планирования экспериментов и формирования выборок; – суть линейной регрессии, ANOVA, логистической и пуассоновской регрессии; – базовые принципы многомерного анализа (PCA, кластерный анализ) и их реализацию в R. уметь: – формулировать содержательную гипотезу и переводить её на язык статистической модели; – проводить разведочный анализ данных (EDA) и корректно выбирать метод в зависимости от типа переменных; – строить и интерпретировать простые регрессионные и ANOVA-модели, проверять допущения.	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа	знать: – логику проверки статистических гипотез, p-value и его интерпретацию; – основы планирования экспериментов и формирования выборок; – суть линейной регрессии, ANOVA, логистической и пуассоновской регрессии; – базовые принципы многомерного анализа (PCA, кластерный анализ) и их реализацию в R. уметь: – самостоятельно проектировать схему статистического анализа под задачи исследования, обосновывая выбор	Защита практикума	Шкала оценивания защиты практикума.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
			моделей и методов; – строить смешанные модели, выполнять выбор на основе информационных критериев (AIC, BIC), проводить многомерный анализ (RDA, PERMANOVA); – интерпретировать результаты в контексте биологической/экологической задачи, выявлять скрытые закономерности и формулировать практические рекомендации. владеть: – техникой аналитического письма: подготовка раздела «Методы» и «Результаты» в научной статье с использованием современного статистического аппарата; – способностью критически оценивать статистические подходы в чужих публикациях и воспроизводить заявленные результаты.		

Шкала оценивания защиты практической работы(практикума)

Показатель	Баллы
Студент демонстрирует глубокое понимание статистической теории, безошибочно выбирает и обосновывает метод анализа в соответствии с типом данных и поставленной гипотезой. Код на R корректен, хорошо структурирован и документирован (комментарии, понятные имена переменных). Визуализация выполнена на высоком уровне (ggplot2), все графики снабжены информативными подписями. Выводы содержательны, корректно интерпретированы в терминах исходной биологической/экологической задачи, указаны ограничения анализа. При защите студент свободно аргументирует выбор модели, отвечает на дополнительные вопросы, способен модифицировать код под изменённые условия.	9–10
Метод анализа в целом верен, но имеются небольшие неточности в обосновании или выборе дополнительных критериев. Код рабочий, однако стиль оформления или документирование не всегда оптимальны	7–8

Показатель	Баллы
(отсутствуют некоторые комментарии). Графики построены корректно, но могут быть недостаточно информативны (неполные подписи, неоптимальный выбор масштаба или цветовой палитры). Выводы правильные, но не всегда полные — могут быть упущены некоторые важные аспекты интерпретации. При защите студент отвечает на большинство вопросов, демонстрирует понимание материала, хотя некоторые сложные моменты требуют уточнений.	
Студент в целом представляет, какой метод следует применить, но допускает ошибки в реализации или интерпретации. Код выполняется, но может содержать неоптимальные конструкции или ошибки в расчётах, которые студент не может самостоятельно исправить. Графики присутствуют, но слабо отражают суть результатов. Выводы поверхностны или частично не соответствуют полученным результатам. При защите студент отвечает лишь на простые вопросы, путается в терминологии, нуждается в наводящих подсказках.	5–6
Статистический анализ выполнен с грубыми ошибками: неверно выбран метод, нарушены основные допущения, результаты неадекватны. Код содержит синтаксические ошибки или не позволяет получить осмысленный результат. Графики отсутствуют или не соответствуют данным. Выводы либо отсутствуют, либо полностью ошибочны. Студент не может объяснить ни один этап работы, не ориентируется в назначении использованных функций.	3–4
Практическая работа не выполнена, код не представлен, отчёт не сдан. Студент не владеет минимально необходимыми сведениями, не может ответить ни на один вопрос по теме работы.	0–2

Максимальное количество баллов – 50.

Шкала оценивания опроса

Показатель	Баллы
Студент демонстрирует всестороннее, систематическое знание методов статистического анализа экологических и генетических данных. Ответы на вопросы полные, логически выстроенные, подкреплены конкретными примерами из практики. Студент свободно владеет понятийным аппаратом (генеральная совокупность, выборка, p-value, мощность критерия, типы ошибок, AIC, F-статистики и др.), способен сравнивать различные подходы и обосновывать выбор того или иного метода. Ответы аргументированы, при необходимости сопровождаются формулами и фрагментами кода. Речь грамотная, терминология используется точно.	9–10
Студент показывает хорошее знание материала, основные понятия и определения излагает верно. Ответы достаточно полные, но могут содержать незначительные неточности или негрубые ошибки. Студент умеет объяснить логику проверки гипотез, смысл коэффициентов модели, но может испытывать затруднения при объяснении тонких моментов (например,	7–8

Показатель	Баллы
различие между AIC и BIC, интерпретация взаимодействий). На дополнительные вопросы отвечает, однако не всегда столь же уверенно, как на основные.	
Студент имеет общее представление о предмете, но знание фрагментарно и поверхностно. Ответы кратки, неполны, часто требуют наводящих вопросов. Студент может дать определения основным понятиям, но допускает ошибки при их применении к конкретным задачам (например, путает параметрические и непараметрические тесты). Аргументация отсутствует или слабая. Прослеживаются затруднения в установлении логических связей между этапами анализа.	5–6
Студент демонстрирует существенные пробелы в знаниях. Ответы отрывочны, содержат грубые фактические ошибки, основные понятия искажены или не раскрыты. Студент не может объяснить даже базовые принципы статистического вывода (например, что такое нулевая гипотеза, как интерпретировать p-value). На большинство вопросов отвечает неверно или совсем не отвечает, не способен поддержать беседу по теме.	3–4
Д). Студент отказывается отвечать, либо его ответы полностью не соответствуют заданному вопросу, свидетельствуют о полном незнании материала. Не может воспроизвести ни одного определения, не ориентируется в тематике.	0–2

Максимальное количество баллов - 10

Шкала оценивания тестирования

Показатель	Баллы
Студент демонстрирует отличное знание теоретических основ и практических аспектов статистического анализа на R. Правильные ответы составляют не менее 90% от общего числа вопросов теста. Допускается не более 10% ошибочных ответов, которые, как правило, касаются наиболее тонких и специфических вопросов (например, синтаксические особенности редких функций).	9–10
Студент показывает хорошее усвоение материала. Доля правильных ответов находится в диапазоне 70–89%. Возможны отдельные ошибки в вопросах, касающихся деталей реализации методов (конкретные параметры функций пакетов <code>vegan</code> , <code>adeonet</code>) или нюансов интерпретации диагностических графиков, но общее понимание предмета твёрдое.	7–8
Студент демонстрирует удовлетворительное знание базовых разделов. Доля правильных ответов составляет 50–69%. Имеются пробелы в знании некоторых тем (например, F-статистики, пермутационные тесты), а также ошибки в интерпретации ключевых понятий (p-value, доверительный интервал). Тем не менее, студент справляется с простыми вопросами,	5–6

Показатель	Баллы
показывая минимально необходимый уровень компетенции.	
Студент показывает слабое знание материала. Правильных ответов от 30% до 49%. Большинство вопросов выполнено неверно, либо ответы выбраны наугад. Студент путает базовые определения, не способен различить типы переменных, не знает назначения основных статистических тестов. Тестирование выявляет серьёзные пробелы в подготовке.	3–4
Д). Доля правильных ответов менее 30%. Ответы практически полностью ошибочны, либо тест не пройден. Студент не владеет даже элементарными знаниями по дисциплине, не ориентируется в её содержании.	0–2

Максимальное количество баллов - 20.

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Примерные темы практических работ

Практическая работа № 1

Тема: Описательная статистика, визуализация данных и сравнение двух групп

Цель: Освоить базовые операции с данными в R: импорт, очистку, расчёт описательных статистик и визуализацию. Научиться выбирать и проводить параметрические (t-тест) и непараметрические (тест Манна–Уитни) тесты для сравнения двух независимых выборок, интерпретировать результаты.

Оборудование и программное обеспечение: компьютерный класс, R (версия 4.0+), RStudio Desktop, пакеты readxl, dplyr, ggplot2, car, rstatix.

Ход работы:

- Импорт и первичный осмотр данных.** Скачайте тестовый набор данных (например, масса тела двух популяций ящериц) в формате CSV. Импортируйте командой `data <- read.csv("lizards.csv")`. Изучите структуру: `str(data)`, `head(data)`, `summary(data)`. Проверьте наличие пропусков: `colSums(is.na(data))`. При необходимости замените или удалите NA, используя `na.omit()` или `tidyr::replace_na()`.
- Описательные статистики по группам.** Сгруппируйте данные по популяции (`group_by(data, Population)`) и вычислите: среднее, стандартное отклонение, медиану, минимум, максимум, объём выборки — с помощью `summarise()`. Результат выведите на экран.
- Визуализация.** Постройте гистограммы для каждой группы (`ggplot + geom_histogram`), наложите оценку плотности. Создайте боксплоты для сравнения распределений (`geom_boxplot`). Оцените визуально симметрию, наличие выбросов, различие медиан.
- Проверка нормальности.** Для каждой группы выполните тест Шапиро–Уилка: `shapiro.test(data$Mass[data$Population == "A"])`. Постройте Q-Q графики (`qqnorm`, `qqline`). Сделайте вывод о возможности применения параметрического теста.
- Проверка гомогенности дисперсий.** Примените тест Левена: `leveneTest(Mass ~ Population, data = data)` из пакета `car`. Если $p > 0.05$, дисперсии можно считать равными.
- Сравнение групп.** Если условия соблюдены, выполните t-тест Стьюдента для независимых выборок: `t.test(Mass ~ Population, data = data, var.equal = TRUE)`. В противном случае — тест Манна–Уитни: `wilcox.test(Mass ~ Population, data = data)`. Запишите значение статистики, p-

уровень, 95% доверительный интервал разности средних.

7. **Интерпретация.** Сформулируйте нулевую и альтернативную гипотезы. Опишите, есть ли статистически значимые различия между популяциями. Обсудите возможные биологические причины различий.
8. **Оформление отчёта.** Создайте R Markdown документ, включите код, выводы, графики. Экспортируйте в HTML или PDF.

Практическая работа № 2

Тема: Линейная регрессия и дисперсионный анализ (ANOVA)

Цель: Освоить построение и диагностику линейных моделей средствами R. Научиться проводить простую и множественную линейную регрессию, однофакторный и двухфакторный дисперсионный анализ, интерпретировать коэффициенты и проверять допущения модели.

Оборудование и программное обеспечение: компьютерный класс, R, RStudio, пакеты ggplot2, car, emmeans, performance.

Ход работы:

1. **Загрузка данных.** Используйте встроенный набор mtcars или экологический датасет (например, зависимость обилия вида от температуры и осадков). Импортируйте и осмотрите: `?mtcars`, `str(mtcars)`.
2. **Простая линейная регрессия.** Постройте модель зависимости расхода топлива (mpg) от мощности (hp): `model1 <- lm(mpg ~ hp, data = mtcars)`. Выведите сводку: `summary(model1)`. Запишите уравнение регрессии, значение R^2 , p-уровень для коэффициента наклона. Интерпретируйте коэффициент: на сколько в среднем изменяется mpg при увеличении hp на единицу.
3. **Визуализация.** Постройте диаграмму рассеяния с линией регрессии: `ggplot(mtcars, aes(hp, mpg)) + geom_point() + geom_smooth(method = "lm")`. Добавьте доверительную полосу.
4. **Диагностика остатков.** Проверьте допущения: нормальность остатков (`shapiro.test(resid(model1))`), график остатков от предсказанных значений (`plot(model1, which = 1)`), график Q-Q (`plot(model1, which = 2)`). С помощью `check_model(model1)` из пакета performance получите сводный диагностический отчёт. Обнаружьте возможные выбросы и влиятельные наблюдения.
5. **Множественная регрессия.** Добавьте в модель вес автомобиля (wt) и число цилиндров (cyl): `model2 <- lm(mpg ~ hp + wt + cyl, data = mtcars)`. Сравните модели по скорректированному R^2 и AIC (`AIC(model1, model2)`). Интерпретируйте коэффициенты при новых предикторах.
6. **Однофакторный ANOVA.** Рассмотрите фактор числа цилиндров (cyl) как категориальный. Выполните ANOVA: `aov1 <- aov(mpg ~ factor(cyl), data = mtcars)`, `summary(aov1)`. При значимом результате выполните post-hoc тест Тьюки: `TukeyHSD(aov1)`. Какие группы различаются значимо?
7. **Двухфакторный ANOVA.** Постройте модель с двумя факторами и их взаимодействием: `aov2 <- aov(mpg ~ factor(cyl) * factor(vs), data = mtcars)`. Проинтерпретируйте таблицу ANOVA. Если взаимодействие значимо, проанализируйте простые эффекты с помощью `emmeans`.
8. **Выводы.** Опишите, какие факторы значимо влияют на расход топлива, какова доля объяснённой дисперсии. Обсудите биологический (экологический) смысл полученных зависимостей, если бы это были экологические данные.

Практическая работа № 3

Тема: Обобщённые линейные модели (GLM): логистическая и пуассоновская регрессия

Цель: Научиться строить и интерпретировать обобщённые линейные модели для бинарных и счётных данных. Освоить выбор функции связи, оценку качества модели и визуализацию предсказаний.

Оборудование и программное обеспечение: компьютерный класс, R, RStudio, пакеты ggplot2, dplyr, pROC (для ROC-кривой), MASS.

Ход работы:

1. **Логистическая регрессия: подготовка.** Загрузите данные о выживаемости организмов в зависимости от дозы токсиканта (датасет `dose_response.csv`).
Переменные: `dose` (числовая), `survived` (0/1). Осмотрите данные. Постройте scatter plot с точками разного цвета для выживших/погибших.
2. **Построение модели.** Выполните логистическую регрессию: `glm1 <- glm(survived ~ dose, data = data, family = binomial)`. Выведите `summary(glm1)`. Интерпретируйте коэффициенты: что означает экспонента коэффициента (odds ratio)? Является ли доза значимым предиктором?
3. **Предсказание и визуализация.** Создайте вектор новых значений дозы: `newdata <- data.frame(dose = seq(min(data$dose), max(data$dose), length.out = 100))`. Получите предсказанные вероятности: `newdata$prob <- predict(glm1, newdata, type = "response")`. Постройте график: исходные точки и сигмоидная кривая.
4. **Оценка качества.** Постройте ROC-кривую с помощью пакета `rROC`: `roc_obj <- roc(data$survived, fitted(glm1))`, `plot(roc_obj)`. Вычислите AUC. Определите оптимальный порог отсечения по индексу Йодена.
5. **Пуассоновская регрессия: данные.** Загрузите датасет о численности паразитов на особях хозяина в зависимости от возраста и пола.
Переменные: `parasites` (счёт), `age` (числовая), `sex` (фактор). Постройте гистограмму распределения числа паразитов. Отметьте, что данные счётные и неотрицательные.
6. **Построение модели.** Постройте пуассоновскую GLM: `glm2 <- glm(parasites ~ age + sex, data = data, family = poisson)`. Проверьте наличие сверхрассеяния (overdispersion): разделите остаточный девианс на число степеней свободы. Если $>>1$, используйте отрицательную биномиальную модель (`glm.nb` из пакета `MASS`). Сравните модели по AIC.
7. **Интерпретация.** Для выбранной модели интерпретируйте коэффициенты: как возраст и пол влияют на ожидаемое число паразитов? Постройте график предсказанных значений вместе с доверительными интервалами. Опишите результаты в экологическом контексте.
8. **Выводы.** Сравните, в каких ситуациях применяются логистическая и пуассоновская регрессии, как интерпретировать их коэффициенты. Оформите отчёт.

Практическая работа № 4

Тема: Многомерный анализ экологических сообществ: ординация, кластеризация и оценка разнообразия

Цель: Освоить методы ординации (PCA, NMDS) и кластерного анализа для описания структуры экологических сообществ. Научиться вычислять и интерпретировать индексы альфа- и бета-разнообразия, визуализировать многомерные данные с помощью пакета `vegan`.

Оборудование и программное обеспечение: компьютерный класс, R, RStudio, пакеты `vegan`, `ggplot2`, `factoextra`, `viridis`.

Ход работы:

1. **Подготовка данных.** Импортируйте матрицу «виды × площадки» (датасет `community.csv`) и таблицу с переменными среды (`env.csv`). Осмотрите данные. Рассчитайте общее обилие на каждой площадке, число видов.
2. **Индексы альфа-разнообразия.** Для каждой площадки вычислите: видовое богатство (`specnumber()`), индекс Шеннона (`diversity(comm, index = "shannon")`), индекс Симпсона (`diversity(comm, index = "simpson")`). Постройте боксплоты или гистограммы распределения индексов. Свяжите разнообразие с градиентами среды, используя корреляцию или регрессию.
3. **Ординация PCA.** Выполните PCA по видовой матрице (после трансформации, например, Хеллингера: `decostand(comm, "hellinger")`). Используйте `rda()` из `vegan`. Постройте ординационную диаграмму: `plot(pca)`, добавьте подписи площадок. Визуализируйте в `ggplot2` с эллипсами групп, если есть категориальная переменная. Определите, какие виды вносят наибольший вклад в главные компоненты.
4. **NMDS.** Постройте неметрическое многомерное шкалирование: `nmfs <- metaMDS(comm,`

distance = "bray", k = 2). Оцените стресс (nmds\$stress). Если стресс >0.2, попробуйте увеличить число осей до 3. Постройте ординационную диаграмму с помощью ordiplot(nmds, type = "t"). Наложите векторы экологических переменных функцией envfit(nmds, env) и визуализируйте значимые ($p < 0.05$).

5. **Кластерный анализ.** Постройте иерархическую кластеризацию площадок на основе расстояния Брэя–Кёртиса: `d <- vegdist(comm, "bray")`, `clust <- hclust(d, method = "ward.D2")`. Визуализируйте дендрограмму (`plot(clust)`). Добавьте прямоугольники, выделяющие группы (`rect.hclust(clust, k = 3)`). Сравните группы с результатами NMDS.
6. **Интерпретация.** Опишите, какие факторы среды сильнее всего связаны с вариацией в составе сообществ. Какие виды характерны для разных условий? Какие площадки наиболее сходны?
7. **Отчёт.** Создайте сводный документ с графиками PCA, NMDS, дендрограммой и выводами о структуре сообщества.

Практическая работа № 5

Тема: Популяционно-генетический анализ: частоты аллелей, равновесие Харди–Вайнберга и F-статистики

Цель: Освоить базовые операции с генетическими данными в R: вычисление частот аллелей и генотипов, тестирование отклонений от равновесия Харди–Вайнберга, расчёт и интерпретацию F-статистик Райта. Визуализировать генетическую структуру популяций с помощью пакетов `adegenet` и `hierfstat`.

Оборудование и программное обеспечение: компьютерный класс, R, RStudio, пакеты `adegenet`, `hierfstat`, `pegas`, `poppr`, `ggplot2`.

Ход работы:

1. **Загрузка и конвертация данных.** Скачайте файл в формате Genepop (`fish.gen`) или используйте встроенный набор `nancycats` (пакет `adegenet`). Импортируйте данные: `data(nancycats)`, `genind_obj <- nancycats`. Изучите объект: `genind_obj`. Проверьте число популяций, локусов, особей.
2. **Частоты аллелей.** Рассчитайте частоты аллелей для каждой популяции: `allele_freq <- allele(genind_obj)`. Визуализируйте распределение частот для нескольких локусов с помощью столбчатых диаграмм. Вычислите наблюдаемую и ожидаемую гетерозиготность: `ho <- summary(genind_obj)$Hobs`, `he <- summary(genind_obj)$Hexp`. Постройте график сравнения H_o и H_e .
3. **Тест Харди–Вайнберга.** Проведите тест Харди–Вайнберга для каждой популяции по каждому локусу с помощью `hw.test(genind_obj, B = 1000)` из пакета `pegas` (или `hwe.test` из `adegenet`). Интерпретируйте результаты: в каких популяциях и локусах наблюдаются значимые отклонения? Возможные причины (инбридинг, нуль-аллели, эффект Вальунда).
4. **Оценка F-статистик.** Конвертируйте объект `genind` в `hierfstat` формат: `hf <- genind2hierfstat(genind_obj)`. Рассчитайте F-статистики по всем локусам: `fstat <- pp.fst(hf)`. Определите F_{it} , F_{st} , F_{is} . Интерпретируйте F_{st} : насколько велика дифференциация между популяциями (шкала Райта)? Проведите попарные сравнения F_{st} : `pairwise.fst(hf)`. Постройте тепловую карту попарных F_{st} .
5. **Визуализация генетической структуры.** Выполните анализ главных компонент (PCA) по частотам аллелей: `pca_obj <- glPca(genind_obj)`. Постройте облако точек, раскрашенное по популяциям. Альтернативно, проведите дискриминантный анализ главных компонент (DAPC): `dapc_obj <- dapc(genind_obj, n.pca = 20, n.da = 3)`, `scatter(dapc_obj)`. Интерпретируйте, насколько хорошо разделяются популяции.
6. **Выводы.** Опишите уровень генетического разнообразия в популяциях, наличие инбридинга, степень дифференциации. Предположите, какие эволюционные или экологические факторы могли привести к наблюдаемой структуре.
7. **Отчёт.** Оформите результаты в виде аналитического отчёта с таблицами частот, графиками PCA/DAPC, тепловой картой F_{st} и содержательной интерпретацией.

Примерные вопросы для защиты практических работ

Практическая работа № 1. Описательная статистика и сравнение групп

1. Какие меры центральной тенденции и разброса вы знаете? В каких ситуациях предпочтительнее использовать медиану вместо среднего?
2. Что показывает p -уровень значимости в статистическом тесте? Как его интерпретируют в контексте нулевой гипотезы?
3. Сформулируйте условия применимости t -критерия Стьюдента для двух независимых выборок. Как проверяют каждое из них в R?
4. В чём принципиальное различие между параметрическими и непараметрическими тестами? Приведите пример, когда использование непараметрического теста обязательно.
5. Как с помощью боксплота оценить наличие различий между группами и выявить возможные выбросы? Какие элементы входят в классический боксплот?

Практическая работа № 2. Линейная регрессия и ANOVA

1. Запишите уравнение простой линейной регрессии. Как интерпретируется коэффициент наклона и свободный член в экологическом контексте?
2. Что показывает коэффициент детерминации R^2 ? В чём его ограничения и почему для сравнения моделей часто используют скорректированный R^2 ?
3. Какие основные допущения лежат в основе линейной регрессионной модели? Как диагностировать их нарушение в R (графически и тестами)?
4. Объясните суть однофакторного дисперсионного анализа. Почему нельзя просто сравнить все пары групп с помощью множества t -тестов?
5. Что такое взаимодействие факторов в двухфакторном ANOVA? Как его наличие влияет на интерпретацию главных эффектов?

Практическая работа № 3. Обобщённые линейные модели (GLM)

1. Почему для моделирования бинарной переменной нельзя применять обычную линейную регрессию? Какие проблемы при этом возникают?
2. Опишите логит-преобразование в логистической регрессии. Как интерпретируется коэффициент модели в терминах отношения шансов?
3. Что такое функция связи в обобщённых линейных моделях? Приведите примеры для нормального, биномиального и пуассоновского распределений.
4. Для каких типов данных применяют пуассоновскую регрессию? Что такое сверхрасеяние (overdispersion) и как с ним бороться?
5. Как оценить качество бинарной классификационной модели с помощью ROC-кривой? Что характеризует показатель AUC?

Практическая работа № 4. Многомерный анализ сообществ

1. В чём суть метода главных компонент (PCA)? Какая информация содержится в собственных значениях и собственных векторах?
2. Чем отличается неметрическое многомерное шкалирование (NMDS) от PCA? В каких случаях NMDS предпочтительнее?
3. Какие индексы альфа-разнообразия вы знаете? Как они учитывают одновременно богатство и выравненность видов?
4. Что такое стресс в NMDS? Какие значения считаются приемлемыми, и как можно снизить стресс?
5. Объясните принцип иерархической кластеризации. В чём разница между методами одиночной связи и методом Уорда?

Практическая работа № 5. Популяционно-генетический анализ

1. Сформулируйте закон Харди–Вайнберга. Какие условия должны выполняться для равновесия в идеальной популяции?
2. Как рассчитать ожидаемые частоты генотипов по наблюдаемым частотам аллелей? Что означает отклонение от равновесия Харди–Вайнберга?
3. Дайте определения F -статистик Райта (F_{is} , F_{it} , F_{st}). Что измеряет каждая из них на внутри- и

межпопуляционном уровнях?

4. Какие программные пакеты в R используются для работы с генетическими данными? Приведите примеры функций для импорта данных и расчёта Fst.
5. Какую информацию даёт анализ главных компонент (PCA) по аллельным частотам? Как визуализировать популяционную структуру с помощью DAPC?

Примерные вопросы для опроса

Что показывает p-уровень значимости в статистическом тесте? Как его интерпретируют в контексте нулевой гипотезы?

1. Какие меры центральной тенденции и разброса вы знаете? В каких ситуациях предпочтительнее использовать медиану вместо среднего?
2. Какие программные пакеты в R используются для работы с генетическими данными?
3. Как рассчитать ожидаемые частоты генотипов по наблюдаемым частотам аллелей? Что означает отклонение от равновесия Харди–Вайнберга?
4. Что такое стресс в NMDS? Какие значения считаются приемлемыми, и как можно снизить стресс?
5. В чём суть метода главных компонент (PCA)? Какая информация содержится в собственных значениях и собственных векторах?
6. Чем отличается неметрическое многомерное шкалирование (NMDS) от PCA? В каких случаях NMDS предпочтительнее?
7. Что показывает коэффициент детерминации R^2 ? В чём его ограничения и почему для сравнения моделей часто используют скорректированный R^2 ?
8. Какие основные допущения лежат в основе линейной регрессионной модели? Как диагностировать их нарушение в R (графически и тестами)?

Примерные тестовые задания

Вариант 1

1. Какая функция в R используется для расчёта среднего арифметического?
А) median(x)
Б) mean(x)
В) sd(x)
Г) var(x)
2. Какое распределение лежит в основе логистической регрессии?
А) Нормальное
Б) Пуассоновское
В) Биномиальное
Г) Экспоненциальное
3. Какой пакет в R предоставляет наиболее широкий набор функций для многомерного анализа экологических сообществ?
А) ggplot2
Б) dplyr
В) vegan
Г) lme4
4. Если p-уровень значимости меньше 0.05, то нулевая гипотеза:
А) Принимается
Б) Отвергается
В) Не формулируется
Г) Считается доказанной
5. Индекс Шеннона является мерой:
А) Генетического сходства
Б) Видового разнообразия
В) Корреляции
Г) Наследственной изменчивости

Вариант 2

1. Какую команду используют для построения линейной модели в R?
А) `lm()`
Б) `glm()`
В) `aov()`
Г) `t.test()`
2. Какой тип данных обычно анализируют с помощью пуассоновской регрессии?
А) Бинарные (0/1)
Б) Непрерывные положительные
В) Счётные (число событий)
Г) Временные ряды
3. Что показывает F-статистика `Fst`?
А) Инбридинг особи относительно популяции
Б) Инбридинг особи относительно всей выборки
В) Дифференциацию между популяциями
Г) Соответствие равновесию Харди–Вайнберга
4. Для проверки нормальности распределения в R часто применяют тест:
А) Левена
Б) Шапиро–Уилка
В) Манна–Уитни
Г) Тьюки
5. В методе главных компонент первая главная компонента объясняет:
А) Наименьшую долю дисперсии
Б) Наибольшую долю дисперсии
В) Корреляцию между переменными
Г) Кластерную структуру

Примерный перечень вопросов к зачёту

1. Основные типы данных в экологических и генетических исследованиях. Шкалы измерений.
2. Импорт, просмотр и очистка данных в R (форматы CSV, Excel, пропуски, фильтрация).
3. Описательные статистики: среднее, медиана, дисперсия, квантили, интерпретация в R.
4. Визуализация данных с помощью `ggplot2`: гистограммы, боксплоты, диаграммы рассеяния.
5. Понятие статистической гипотезы. Нулевая и альтернативная гипотезы, ошибки I и II рода.
6. Уровень значимости p -value: определение, интерпретация, распространённые заблуждения.
7. Параметрические и непараметрические тесты: отличия, условия применения, примеры.
8. t -критерий Стьюдента для одной выборки и для двух независимых выборок.
9. t -критерий для парных (зависимых) выборок: условия, преимущества перед независимым.
10. Критерий Манна–Уитни: суть, гипотезы, реализация в R.
11. Критерий χ^2 (хи-квадрат) для таблиц сопряжённости: использование, интерпретация остатков.
12. Линейная регрессия: постановка задачи, метод наименьших квадратов, уравнение модели.
13. Коэффициенты регрессии: интерпретация, стандартные ошибки, t -тесты для коэффициентов.
14. Коэффициент детерминации R^2 и скорректированный R^2 : расчёт, смысл, ограничения.
15. Допущения линейной регрессии: способы проверки в R (графики остатков, тесты).
16. Мультиколлинеарность: причины, методы обнаружения (VIF), последствия.
17. Однофакторный дисперсионный анализ (ANOVA): модель, разложение дисперсии, F-тест.
18. Post-hoc тесты в ANOVA (Тьюки, Бонферрони): зачем нужны, как интерпретировать.
19. Двухфакторный ANOVA: главные эффекты и взаимодействие, их графическое представление.
20. Обобщённые линейные модели (GLM): компоненты (случайная, систематическая, функция связи).
21. Логистическая регрессия: уравнение, логит-преобразование, отношение шансов (odds ratio).

22. Оценка качества бинарной классификации: матрица ошибок, чувствительность, специфичность, ROC-кривая, AUC.
23. Пуассоновская регрессия: модель для счётных данных, интерпретация коэффициентов как относительного риска.
24. Сверхрасcеяние в пуассоновской регрессии: диагностика, отрицательная биномиальная модель.
25. Информационные критерии AIC и BIC: назначение, сравнение моделей, правило выбора.
26. Выбор между линейной моделью и GLM: постановка задачи, связь распределения отклика и функции связи.
27. Основные принципы многомерного анализа в экологии: задачи ординации и классификации.
28. Анализ главных компонент (PCA): математическая идея, ковариационная матрица, собственные значения и векторы.
29. Интерпретация результатов PCA: loadings, scores, доля объяснённой дисперсии.
30. Неметрическое многомерное шкалирование (NMDS): особенности, метрики расстояний (Bray–Curtis, Jaccard).
31. Сравнение PCA и NMDS: когда какой метод предпочтительнее.
32. Индексы альфа-разнообразия: видовое богатство, индексы Шеннона, Симпсона, их расчёт в R.
33. Бета-разнообразие и его компоненты: замена видов и вложенность, пакет betapart.
34. Кластерный анализ: меры расстояния между объектами, алгоритмы иерархической кластеризации (метод Уорда).
35. Метод k-средних: идея, преимущества, недостатки, выбор числа кластеров.
36. PERMANOVA (adonis2): проверка гипотез о различии состава сообществ, пермутационный подход.
37. Избыточный анализ (RDA) и канонический анализ соответствия (CCA): модели, связь со средовыми переменными.
38. Основы популяционной генетики: аллели, генотипы, локусы, популяция, выборка.
39. Закон Харди–Вайнберга: формулировка, вывод ожидаемых частот генотипов.
40. Тестирование отклонений от равновесия Харди–Вайнберга: χ^2 -тест, точный тест, причины отклонений (инбридинг, нуль-аллели).
41. Инбридинг и коэффициент F (Fis): определение, связь с гетерозиготностью.
42. F-статистики Райта: Fis, Fit, Fst — их биологический смысл, расчёт, интерпретация.
43. Меры генетической дифференциации: GST, Jost's D, сравнение с Fst.
44. Анализ молекулярной дисперсии (AMOVA): разложение генетической изменчивости, реализация в R.
45. Методы визуализации популяционной структуры: PCA по частотам аллелей, DAPC, STRUCTURE.
46. Понятие об эффективном размере популяции (Ne) и методах его оценки.
47. Пакет adegenet: загрузка данных, основные функции, работа с объектами genind и genpop.
48. Пакет hierfstat: вычисление базовых популяционных параметров, попарные Fst, тесты на нейтральность.
49. Требования к воспроизводимости статистического анализа: документирование кода, R Markdown, управление пакетами (renv).
50. Этические аспекты анализа генетических данных: информированное согласие, конфиденциальность, депонирование данных в открытые базы (GenBank, Dryad).

5.4 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Программа освоения дисциплины предусматривает следующие формы текущего контроля: активно участвовать в опросах, подготовить лабораторный практикум и пройти тестирование. Требования к оформлению и выполнению всех предусмотренных в рабочей

программе дисциплин форм отчетности и критериев оценивания отражены в методических рекомендациях.

Максимальное количество баллов, которое может набрать студент в течение семестра за различные виды работ – 80 баллов.

Максимальная сумма баллов, которые может получить студент на зачете – 20 баллов.

Максимальная сумма баллов студентов по изучаемой дисциплине составляет 100 баллов. Формой промежуточной аттестации является зачет, который проходит в форме устного собеседования по вопросам

Шкала оценивания зачета

Показатель	Баллы
Отлично (высокий уровень). Студент демонстрирует всестороннее, глубокое и систематическое знание теоретических основ статистического анализа, свободно ориентируется в современных пакетах R. Ответы на теоретические вопросы полные, логически безупречные, подкреплены формулами, примерами из экологии и популяционной генетики, а также критическим сравнением методов. Практическое задание выполнено безошибочно: корректно выбран метод, код на R эффективен, хорошо документирован, результаты визуализированы на высоком уровне, интерпретация точна и полна. Студент способен объяснить допущения модели, предложить альтернативные подходы и модифицировать анализ под изменённые условия. Речь уверенная, ответы аргументированы.	17–20
Хорошо (базовый уровень). Студент показывает хорошее знание материала, основные теоретические положения излагает верно. Ответы достаточно полны, но могут содержать незначительные неточности или недостаточную детализацию в сложных вопросах (например, различие информационных критериев, интерпретация взаимодействий). Практическое задание выполнено в целом правильно: метод выбран обоснованно, код работает, однако возможны небольшие погрешности в визуализации или неполнота выводов. Студент отвечает на большинство дополнительных вопросов, демонстрирует понимание сути методов, хотя некоторые тонкие аспекты требуют уточнений.	13–16
Удовлетворительно (пороговый уровень). Студент владеет материалом на базовом уровне, но знания фрагментарны. Ответы кратки, часто требуют наводящих вопросов, содержат отдельные ошибки или неточности в терминологии (путает параметрические и непараметрические тесты, неправильно объясняет p-value). Практическое задание выполнено с ошибками: выбранный метод не вполне адекватен данным, код работает, но интерпретация результатов поверхностна или частично неверна. Студент может объяснить лишь основные этапы анализа, слабо аргументирует выбор методов, испытывает затруднения при обсуждении ограничений.	9–12
Неудовлетворительно (низкий уровень). Студент демонстрирует существенные пробелы в знаниях. Теоретические ответы отрывочны, содержат грубые ошибки, основные понятия (нулевая гипотеза, доверительный интервал, линейная модель) искажены или не раскрыты. Практическое задание либо не выполнено, либо выполнено с принципиальными ошибками: неверно выбран метод, код содержит критические ошибки, выводы отсутствуют или полностью ошибочны. Студент не может ответить на дополнительные вопросы, не ориентируется в содержании дисциплины.	5–8
Крайне неудовлетворительно (отсутствие знаний). Студент не отвечает на теоретические вопросы или ответы полностью не соответствуют теме. Практическое задание не выполнено. Не владеет даже минимальными сведениями, не может воспроизвести ни одного определения, не способен написать ни одной строки корректного	0–4

Показатель	Баллы
кода на R.	

Итоговая шкала выставления оценки по дисциплине

При выставлении итоговой оценки преподавателем учитывается работа магистранта в течение всего срока освоения дисциплины, а также баллы, полученные на промежуточной аттестации.

Баллы, полученные магистрантами в течение освоения дисциплины	Оценка по дисциплине
81-100	Зачтено
61-80	Зачтено
41-60	Зачтено
0-40	Не зачтено

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1 Основная литература

- 1 Statistical Learning in Genetics: An Introduction Using R / D. Sorensen. — 2nd ed. — Cham : Springer, 2025. — 693 p. — (Statistics for Biology and Health). — Текст : электронный. — URL: <https://link.springer.com/book/9783031358531>
- 2 Applied Statistical Modelling for Ecologists: A Practical Guide to Bayesian and Likelihood Inference Using R, JAGS, NIMBLE, Stan and TMB / M. Kéry, K. E. Pacífico. — Amsterdam : Elsevier, 2024. — 550 p. — Текст : электронный. — URL: <https://www.elsevier.com/books/applied-statistical-modelling-for-ecologists/kery/978-0-443-13716-7>
- 3 Population Genetics with R: An Introduction for Life Scientists / Á. J. Láruson, F. A. Reed. — Oxford : Oxford University Press, 2024. — 224 p. — Текст : электронный. — URL: <https://global.oup.com/academic/product/population-genetics-with-r-9780198829539>
- 4 Статистический анализ биологических данных с использованием R : учебное пособие / В. К. Шитиков, Г. С. Розенберг. — 3-е изд., перераб. и доп. — Москва : ИНФРА-М, 2025. — 354 с. — Текст : электронный. — URL: <https://znanium.ru/catalog/product/2178651>

6.2 Дополнительная литература

- 1 The Analysis of Ecological Data Using R / T. C. Edwards, D. R. Cutler. — 2nd ed. — Boca Raton : CRC Press, 2024. — 416 с. — Текст : электронный. — URL: <https://www.routledge.com/The-Analysis-of-Ecological-Data-Using-R/Edwards-Cutler/p/book/9781584885962>
- 2 Теория вероятностей в примерах и задачах на языке R : учебник / С. Я. Криволапов. — Москва : ИНФРА-М, 2024. — 412 с. — Текст : электронный. — URL: <https://znanium.ru/catalog/product/2175041>
- 3 Математическая статистика. Практикум : учебное пособие / Т. Г. Апалькова, В. И. Глебов, С. А. Зададаев [и др.]. — Москва : ИНФРА-М, 2025. — 312 с. — Текст : электронный. — URL: <https://znanium.ru/catalog/product/2182536>
- 4 Quantitative Biosciences Companion in R: Dynamics Across Cells, Organisms, and Populations / J. P. O'Dwyer, R. Martinez-Garcia. — Princeton : Princeton University Press, 2024. — 272 с. — Текст : электронный. — URL: <https://press.princeton.edu/books/paperback/9780691256366/quantitative-biosciences-companion-in-r>
- 5 Statistical Methods in Human Genetics / I. R. Konappan. — Singapore : Springer, 2024. — 318 с. — Текст : электронный. — URL: <https://link.springer.com/book/9789819932221>
- 6 Applied Statistics with R: A Practical Guide for the Life Sciences and Beyond / J. C. Touchon. —

- Oxford : Oxford University Press, 2025. — 478 с. — Текст : электронный. — URL: <https://global.oup.com/academic/product/applied-statistics-with-r-9780198873426>
- 7 Multivariate Statistics and Machine Learning in R for Beginners / S. Zamir. — New York : Apress, 2025. — 410 с. — Текст : электронный. — URL: <https://link.springer.com/book/9798868808735>
- 8 Анализ данных с использованием языка программирования R. Часть 1 : учебное пособие / А. В. Малов [и др.]. — Казань : КФУ, 2024. — 156 с. — Текст : электронный. — URL: https://kpfu.ru/portal/docs/F_1245467499/Posobie_R.pdf
- 9 Лабораторный практикум по статистическому анализу экологических и генетических данных в среде R : учебно-методическое пособие / Е. Н. Иванова, А. А. Петров. — Санкт-Петербург : СПбГУ, 2024. — 188 с. — Текст : непосредственный.

6.3 Ресурсы информационно-телекоммуникационной сети «Интернет»

1. The Comprehensive R Archive Network (CRAN) – официальный репозиторий пакетов R и документации. – URL: <https://cran.r-project.org>
2. Bioconductor – набор пакетов R для анализа геномных и омиксных данных. – URL: <https://www.bioconductor.org>
3. RStudio / Posit – интегрированная среда разработки на R, учебные материалы. – URL: <https://posit.co>
4. R-bloggers – агрегатор новостей и учебных статей по статистике на R. – URL: <https://www.r-bloggers.com>
5. GitHub – репозитории с кодом и учебными проектами в области экологической и популяционной генетики на R. – URL: <https://github.com>
6. Stack Overflow – раздел по R и статистике, обсуждение проблем кодирования и интерпретации результатов. – URL: <https://stackoverflow.com/questions/tagged/r>
7. Платформа образовательных курсов DataCamp – онлайн-курсы по статистике на R для экологов и генетиков. – URL: <https://www.datacamp.com>
8. Национальный центр биотехнологической информации США (NCBI) – базы данных генетической и экологической информации, инструменты анализа. – URL: <https://www.ncbi.nlm.nih.gov>
9. Европейский институт биоинформатики (EMBL-EBI) – образовательные ресурсы и инструменты для биоинформатики на R. – URL: <https://www.ebi.ac.uk>
10. Портал «Биоинформатический портал» (bioinf.me) – курсы и ресурсы по анализу генетических данных в R. – URL: <https://bioinf.me>
11. Популяционно-генетическая лаборатория ВШЭ (Москва) – учебные материалы и базы данных по эколого-генетическим системам. – URL: <https://www.hse.ru/lab/popgen>
12. Лаборатория вычислительной биологии и биоинформатики СПбГУ – методические материалы по статистическому анализу биосистем в R. – URL: <https://biocomp.bio.spbu.ru>
13. Кафедра генетики биологического факультета МГУ им. М. В. Ломоносова – учебно-научные материалы по популяционной генетике. – URL: <http://www.genebee.msu.ru>
14. Институт экологии растений и животных УрО РАН (Екатеринбург) – публикации и статистические базы по эколого-генетическим исследованиям. – URL: <https://ipae.uran.ru>
15. Факультет биоинженерии и биоинформатики МГУ – учебные курсы и методические разработки. – URL: <https://fbb.msu.ru>

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

1. Методические рекомендации по организации самостоятельной работы магистров

8. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

Лицензионное программное обеспечение:

Microsoft Windows (версия 10/11 или выше)
Microsoft Office (в составе Word, Excel, PowerPoint)
Kaspersky Endpoint Security для рабочих станций
MATLAB (с пакетами Bioinformatics Toolbox, Statistics and Machine Learning Toolbox) – образовательная лицензия
ПО для удалённого доступа к высокопроизводительным вычислительным кластерам (лицензия на клиентские приложения, например, MobaXterm Professional)

Информационные справочные системы:

Система «КонсультантПлюс»
Специализированная справочная система по биоинформатике – онлайн-документация и форумы Biostars (<https://www.biostars.org>)

Профессиональные базы данных:

fgosvo.ru – Портал Федеральных государственных образовательных стандартов высшего образования
pravo.gov.ru – Официальный интернет-портал правовой информации
www.edu.ru – Федеральный портал Российское образование

Биоинформатические профессиональные базы данных:

NCBI (National Center for Biotechnology Information) – <https://www.ncbi.nlm.nih.gov> (GenBank, RefSeq, SRA, dbSNP, PubMed)
EMBL-EBI (European Bioinformatics Institute) – <https://www.ebi.ac.uk> (ENA, ArrayExpress, PRIDE, InterPro)
Ensembl Genome Browser – <https://www.ensembl.org>
UCSC Genome Browser – <https://genome.ucsc.edu>
UniProt (Universal Protein Resource) – <https://www.uniprot.org>
KEGG (Kyoto Encyclopedia of Genes and Genomes) – <https://www.genome.jp/kegg/>
Gene Ontology (GO) – <http://geneontology.org>
Protein Data Bank (PDB) – <https://www.rcsb.org>

Свободно распространяемое программное обеспечение, в том числе отечественного производства

Системное и общего назначения:

ОМС Плеер (для воспроизведения Электронных Учебных Модулей)
7-zip (архиватор)
Google Chrome (интернет-браузер)
LibreOffice (офисный пакет, альтернатива Microsoft Office)

Виртуализация и контейнеризация:

Oracle VirtualBox (для развёртывания виртуальных машин Linux)
Docker Desktop (для работы с биоинформатическими контейнерами)
WSL (Windows Subsystem for Linux) – встроенный компонент Windows

Операционные системы (свободные версии):

Ubuntu Linux (или аналоги: Debian, CentOS Stream)
Astra Linux Common Edition (отечественная ОС для рабочих станций)

Языки программирования и среды разработки:

Python (интерпретатор) с менеджером пакетов conda (Miniconda3)
R (среда статистической обработки данных)

RStudio Desktop (интегрированная среда для R)
Jupyter Notebook / JupyterLab
Visual Studio Code (редактор кода)

Специализированные биоинформатические пакеты (устанавливаются через conda или из исходных кодов):

FastQC, MultiQC (контроль качества NGS-данных)
Trimmomatic, cutadapt (тримминг прочтений)
BWA, Bowtie2, STAR, HISAT2 (картирование коротких и сплайсированных прочтений)
SAMtools, BEDTools (работа с форматами SAM/BAM, BED)
BCFtools, GATK (поиск и фильтрация генетических вариантов)
SnpEff, Ensembl VEP (аннотация вариантов)
featureCounts, HTSeq-count, Salmon (количественная оценка экспрессии)
DESeq2, edgeR (анализ дифференциальной экспрессии, пакеты R)
MACS2 (поиск пиков ChIP-seq/ATAC-seq)
HOMER, MEME Suite (мотивный анализ)
SPAdes, Canu (сборка геномов de novo)
Bismark, bwa-meth (анализ бисульфитного секвенирования)
Snakemake, Nextflow (менеджеры пайплайнов)

Интегративные платформы и браузеры (веб-интерфейсы, не требующие установки):

Galaxy (<https://usegalaxy.org>) – графическая платформа для построения биоинформатических конвейеров
Integrative Genomics Viewer (IGV) – локальное приложение для визуализации геномных данных

Отечественное свободное ПО:

Язык программирования lScript (при изучении автоматизации, опционально)
Платформа «Биоинформатический портал» (<https://bioinf.me>) – образовательные материалы и облачные вычисления (русская разработка)

Все перечисленные свободно распространяемые программы могут быть установлены как на рабочие станции учащихся, так и использоваться в виртуализированной среде под управлением Linux. При необходимости все пакеты актуализируются и доустанавливаются администратором учебных классов или удалённых серверов.

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Материально-техническое обеспечение дисциплины включает в себя:

- учебные аудитории для проведения лекционных и семинарских занятий, оснащённые учебной мебелью, доской, демонстрационным оборудованием, персональным компьютером преподавателя с подключением к сети «Интернет», мультимедийным проектором и экраном;
- компьютерный класс (лаборатория) для проведения практических занятий, оснащённый рабочими станциями учащихся с установленным специализированным свободно распространяемым и лицензионным программным обеспечением (операционные системы Linux и/или Windows, биоинформатические пакеты, языки программирования R и Python, среды разработки, доступ к менеджерам пакетов conda и контейнеризации), подключением к локальной сети и сети «Интернет»; при необходимости — доступом к высокопроизводительному вычислительному серверу (HPC-кластеру) или облачной вычислительной инфраструктуре;
- помещения для самостоятельной работы обучающихся, оснащённые компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде, профессиональным биоинформатическим базам данных и вычислительным ресурсам
- помещение для самостоятельной работы для инвалидов и лиц с ограниченными возможностями здоровья, оснащенное компьютерной техникой, подключенной к сети Интернет, обеспечено доступом к электронно-образовательной среде Университета, Комплект учебной мебели, персональные компьютеры с подключением к сети Интернет и обеспечением доступа к электронным библиотекам и в электронную информационно-образовательную среду Университета, доска, проектор подвесной.