

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Наумова Наталия Александровна
Должность: Ректор
Дата подписания: 17.08.2026 15:03:23
Уникальный программный ключ:
6b5279da4e034bffa679172803da5b7b559fc69e2

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

УТВЕРЖДЁН

на заседании кафедры общей биологии и
биоэкологии

Протокол от «28» апреля 2026 г. № 9

Заведующий кафедрой

 Гордеев М.И./

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

По дисциплине

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 мониторинг

Квалификация

Магистр

Форма обучения

Очная

Москва
2026

Авторы-составители:

Гордеев М.И., д.б.н., профессор кафедры общей биологии и биоэкологии ФГАОУ ВО «ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»

Москаев А.В., к.б.н., доцент кафедры общей биологии и биоэкологии ФГАОУ ВО «ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»

Темников А.А., старший преподаватель кафедры общей биологии и биоэкологии ФГАОУ ВО «ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»

Фонд оценочных средств по дисциплине «Основы математической статистики эколого-генетических систем на языке программирования R» составлен в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 06.04.01 Биология, утверждённого приказом МИНОБРНАУКИ России от 11.08.2020 № 934.

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

Оглавление

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы	4
2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания	4
3. Типовые контрольные задания	8
3.1. Тестовые задания	8
3.2. Ситуационные задачи (краткий ответ)	10
3.3. Открытые вопросы (краткий ответ)	10
3.4. Задания для практических работ	11
3.5. Вопросы для опроса и собеседования	12
3.6. Вопросы к зачёту	14
4. Методические материалы, определяющие процедуры оценивания	15
4.1. Общие положения	15
4.2. Формы текущего контроля и их балльная оценка	15
4.3. Процедура проведения зачёта	16
4.4. Порядок ликвидации академической задолженности	17

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 (лекции, практические занятия). 2. Самостоятельная работа (подготовка к опросам, выполнение практических работ).
СПК-4. Способен проводить экспертно-аналитическую работу при проведении научных исследований и экспериментальных работ.	1. Работа на учебных занятиях (лекции, практические занятия). 2. Самостоятельная работа (подготовка к опросам, выполнение практических работ).

2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
СПК-2 – Способен использовать программы для эпигенетического анализа и качества эпигенетических данных	Пороговый	1. Работа на учебных занятиях	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, критерий χ^2).	Опрос, тестирование	Шкала оценивания опроса. Шкала оценивания тестирования.
	Продвинутый	1. Работа на	уметь: – создавать воспроизводимые отчёты (R Markdown) с включением кода, результатов и интер-	Защита практикума	Шкала оценивания защиты практикума.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
		учебных занятиях 2. Самостоятельная работа	претации; – применять продвину- тые графические системы (ggplot2 с настройкой тем, facets) для визу- ализации многомерных данных; – автоматизировать рутинные опе- рации с помощью собственных функций и циклов; – критически оценивать качество исходных эпигенетических/экологических данных (проверка на артефакты, автокорреляцию) и выбирать под- ходящий статистический инстру- ментарий. владеть: – навыками работы с пакетами для контроля качества и препроцессинга (dplyr, tidyr, janitor); – техникой состав- ления аналитических скриптов, обеспечивающих полную воспро- изводимость от исходных файлов до финальных выводов.		
СПК-4 – Способен проводить экспертно- аналитиче- скую работу при проведе- нии научных исследова- ний и экспе- рименталь- ных работ	Поро- говый	1. Ра- бота на учеб- ных заня- тиях	знать: – логику проверки стати- стических гипотез, p-value и его интерпретацию; – основы плани- рования экспериментов и форми- рования выборок; – суть линейной регрессии, ANOVA, логистиче- ской и пуассоновской регрессии; – базовые принципы многомер- ного анализа (PCA, кластерный анализ) и их реализацию в R. уметь: – формулировать содер- жательную гипотезу и переводить её на язык статистической мо- дели; – проводить разведочный анализ данных (EDA) и корректно выбирать метод в зависимости от типа переменных; – строить и ин- терпретировать простые регресси- онные и ANOVA-модели, прове- рять допущения.	Опрос, тести- рова- ние	Шкала оце- нивания опроса. Шкала оце- нивания те- стирова- ния.
	Про- двину- тый	1. Ра- бота на учеб- ных	уметь: – самостоятельно проекти- ровать схему статистического ана- лиза под задачи исследования, обосновывая выбор моделей и ме- тодов; – строить смешанные мо- дели, выполнять выбор на основе	Защита прак- тикума	Шкала оце- нивания за- щиты прак- тикума.

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
		занятиях 2. Самостоятельная работа	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 тестирования

Показатель	Баллы
Отлично (высокий уровень). Студент демонстрирует отличное знание теоретических основ и практических аспектов статистического анализа на R. Правильные ответы составляют не менее 90% от общего числа вопросов теста.	9–10
Хорошо (базовый уровень). Доля правильных ответов – 70–89%.	7–8
Удовлетворительно (пороговый уровень). Доля правильных ответов – 50–69%.	5–6
Неудовлетворительно (низкий уровень). Доля правильных ответов – 30–49%.	3–4
Крайне неудовлетворительно. Доля правильных ответов менее 30%.	0–2

Максимальное количество баллов – 20.

Шкала оценивания опроса и собеседования

Показатель	Баллы
Отлично (высокий уровень). Студент демонстрирует всестороннее, систематическое знание методов статистического анализа экологических и генетических данных. Ответы полные, логически выстроенные, подкреплены примерами. Свободно владеет понятийным аппаратом, способен сравнивать подходы, отвечает на дополнительные вопросы.	9–10
Хорошо (базовый уровень). Студент показывает хорошее знание материала, основные понятия излагает верно. Ответы достаточно полные, но могут содержать незначительные неточности. На дополнительные вопросы отвечает, но не всегда уверенно.	7–8
Удовлетворительно (пороговый уровень). Студент имеет общее представление о предмете, но знание фрагментарно. Ответы неполны, требуют наводящих вопросов. Допускает ошибки при применении понятий к конкретным задачам.	5–6
Неудовлетворительно (низкий уровень). Ответы содержат грубые ошибки, основные понятия искажены или не раскрыты. На большинство вопросов отвечает неверно.	3–4
Крайне неудовлетворительно. Отказывается отвечать или ответы полностью не соответствуют заданию.	0–2

Максимальное количество баллов – 10.

Шкала оценивания защиты практической работы

Показатель	Баллы
Отлично (высокий уровень). Студент демонстрирует глубокое понимание статистической теории, безошибочно выбирает и обосновывает метод анализа. Код на R корректен, хорошо структурирован и документирован. Визуализация выполнена на высоком уровне. Выводы содержательны, корректно интерпретированы. При защите свободно аргументирует выбор модели, отвечает на дополнительные вопросы.	9–10
Хорошо (базовый уровень). Метод анализа в целом верен, но имеются небольшие неточности в обосновании. Код рабочий, но стиль оформления или документирование не всегда оптимальны. Графики построены корректно, но могут быть недостаточно информативны. Выводы правильные, но не всегда полные. При защите отвечает на большинство вопросов.	7–8
Удовлетворительно (пороговый уровень). Студент в целом представляет, какой метод следует применить, но допускает ошибки в реализации или интерпретации. Код выполняется, но может содержать неоптимальные конструкции. Графики присутствуют, но слабо отражают суть результатов. Выводы поверхностны. При защите отвечает лишь на простые вопросы.	5–6

Показатель	Баллы
Неудовлетворительно (низкий уровень). Статистический анализ выполнен с грубыми ошибками: неверно выбран метод, нарушены допущения. Код содержит синтаксические ошибки. Графики отсутствуют или не соответствуют данным. Выводы отсутствуют или ошибочны. Студент не может объяснить ни один этап работы.	3–4
Крайне неудовлетворительно. Практическая работа не выполнена, код и отчёт не представлены.	0–2

Максимальное количество баллов – 50 (5 работ по 10 баллов).

3. Типовые контрольные задания

3.1. Тестовые задания

Вариант 1

Выберите один правильный ответ из четырёх предложенных.

- Какая функция в R используется для расчёта среднего арифметического?
 - median(x)
 - mean(x)
 - sd(x)
 - var(x)

Ответ: б
- Какое распределение лежит в основе логистической регрессии?
 - нормальное
 - пуассоновское
 - биномиальное
 - экспоненциальное

Ответ: в
- Какой пакет в R предоставляет наиболее широкий набор функций для многомерного анализа экологических сообществ?
 - ggplot2
 - dplyr
 - vegan
 - lme4

Ответ: в
- Если р-уровень значимости меньше 0.05, то нулевая гипотеза:
 - принимается
 - отвергается
 - не формулируется
 - считается доказанной

Ответ: б
- Индекс Шеннона является мерой:
 - генетического сходства
 - видового разнообразия
 - корреляции
 - наследственной изменчивости

Ответ: б
- Какой пакет в R используется для построения обобщённых линейных моделей?
 - stats
 - lme4

- в) MASS
- г) все перечисленные

Ответ: г

7. Что такое AIC?

- а) коэффициент детерминации
- б) информационный критерий Акаике
- в) тест на нормальность
- г) индекс разнообразия

Ответ: б

8. Какой тип данных анализируют с помощью пуассоновской регрессии?

- а) бинарные (0/1)
- б) непрерывные положительные
- в) счётные (число событий)
- г) временные ряды

Ответ: в

9. 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

Ответ: в

10. Для проверки нормальности распределения в R часто применяют тест:

- а) Левена
- б) Шапиро–Уилка
- в) Манна–Уитни
- г) Тьюки

Ответ: б

Выберите несколько правильных ответов из предложенных.

11. Какие функции в R используются для построения линейной модели?

- а) lm()
- б) glm()
- в) aov()
- г) lmer()

Ответ: а, в, г

12. Какие пакеты в R предназначены для популяционно-генетического анализа?

- а) adegenet
- б) hierfstat
- в) vegan
- г) poppr

Ответ: а, б, г

13. Какие методы относятся к методам многомерной ординации?

- а) PCA
- б) NMDS
- в) RDA
- г) ANOVA

Ответ: а, б, в

Установите правильную последовательность.

14. Расположите этапы статистического анализа в правильном порядке:

- 1 – формулировка гипотезы
- 2 – сбор данных
- 3 – импорт и очистка данных
- 4 – выбор метода анализа

5 – интерпретация результатов

Ответ: $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$

15. Установите последовательность проверки допущений линейной регрессии:

1 – нормальность остатков

2 – гомоскедастичность

3 – линейность связи

4 – отсутствие мультиколлинеарности

Ответ: $3 \rightarrow 1 \rightarrow 2 \rightarrow 4$

3.2. Ситуационные задачи (краткий ответ)

1. В R выполнен t-тест: $t.test(x, y, var.equal = TRUE)$. Получено $p = 0.03$. Что следует сделать с нулевой гипотезой при $\alpha = 0.05$?

Ответ: отвергнуть

2. В модели линейной регрессии $R^2 = 0.85$. Какой процент дисперсии объясняется моделью?

Ответ: 85

3. Для пуассоновской регрессии получено отношение остаточного девианса к числу степеней свободы 2.5. О чём это говорит?

Ответ: свёрхраспечение

4. В методе главных компонент первая главная компонента объясняет 60% дисперсии. Какая доля остаётся необъяснённой?

Ответ: 40

5. В популяции частота аллеля $A = 0.7$. Какова ожидаемая частота гетерозигот в равновесной популяции?

Ответ: 0,42

6. В двухфакторном ANOVA обнаружено значимое взаимодействие факторов. Как интерпретировать главные эффекты?

Ответ: осторожно, не разделяя

7. В ROC-анализе получено $AUC = 0.92$. Как оценивается качество модели?

Ответ: отличное

8. $F_{st} = 0.15$ между двумя популяциями. Какой уровень дифференциации это характеризует?

Ответ: умеренный

9. В модели логистической регрессии коэффициент при предикторе равен 1.2. Чему равно отношение шансов (odds ratio)?

Ответ: 3,32

10. В R для оценки качества модели использована функция $AIC()$. Меньшее значение AIC указывает на:

Ответ: лучшую модель

3.3. Открытые вопросы (краткий ответ)

1. Как называется функция в R для построения линейной регрессии?

Ответ: *lm*

2. Какой пакет в R используется для визуализации данных?

Ответ: *ggplot2*

3. Как называется статистический тест для сравнения двух независимых выборок без предположения о нормальности?

Ответ: манна-уитни

4. Какой информационный критерий используется для сравнения моделей?

Ответ: *aic*

5. Как называется метод ординации, использующий собственные значения и векторы ковариационной матрицы?
Ответ: pca
6. Какой показатель используется для оценки генетической дифференциации популяций?
Ответ: fst
7. Как называется метод кластеризации, основанный на последовательном объединении объектов?
Ответ: иерархическая кластеризация
8. Какой тест используется для проверки нормальности распределения в R?
Ответ: шапиро-уилка
9. Как называется модель для бинарных данных в GLM?
Ответ: логистическая регрессия
10. Какой пакет в R используется для анализа популяционно-генетических данных?
Ответ: adegenet

3.4. Задания для практических работ

Практическая работа № 1. 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.

Ход работы:

1. Импорт и первичный осмотр данных (read.csv, str, head, summary).
2. Проверка пропусков (colSums(is.na(data))).
3. Расчёт описательных статистик по группам (group_by, summarise).
4. Визуализация (гистограммы, боксплоты в ggplot2).
5. Проверка нормальности (тест Шапиро–Уилка, Q-Q графики).
6. Проверка гомогенности дисперсий (тест Левена).
7. Сравнение групп: t-тест или тест Манна–Уитни.
8. Интерпретация результатов, формулировка выводов.
9. Оформление отчёта в R Markdown.

Критерии выполнения: код корректен; графики подписаны; выводы обоснованы.

Практическая работа № 2. 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.

Ход работы:

1. Загрузка данных (mtcars или экологический датасет).
2. Простая линейная регрессия (lm, summary). Интерпретация коэффициентов.
3. Визуализация (диаграмма рассеяния с линией регрессии).
4. Диагностика остатков (нормальность, гомоскедастичность, выбросы).
5. Множественная регрессия (добавление предикторов). Сравнение моделей по R^2 и AIC.
6. Однофакторный ANOVA (aov, TukeyHSD).
7. Двухфакторный ANOVA (главные эффекты и взаимодействие).
8. Выводы и интерпретация.

Практическая работа № 3. Обобщённые линейные модели (GLM): логистическая и пуассоновская регрессия

Цель: Научиться строить и интерпретировать обобщённые линейные модели для бинарных и счётных данных.

Ход работы:

1. Логистическая регрессия: загрузка данных (выживаемость в зависимости от дозы).
2. Построение модели (glm, family = binomial). Интерпретация коэффициентов (odds ratio).
3. Визуализация сигмоидной кривой.
4. Оценка качества (ROC-кривая, AUC).
5. Пуассоновская регрессия: загрузка счётных данных (число паразитов).
6. Построение модели (glm, family = poisson). Проверка сверхрасеяния.
7. При необходимости — отрицательная биномиальная модель (glm.nb).
8. Интерпретация коэффициентов, визуализация предсказаний.
9. Выводы.

Практическая работа № 4. Многомерный анализ экологических сообществ: ординация, кластеризация и оценка разнообразия

Цель: Освоить методы ординации (PCA, NMDS) и кластерного анализа для описания структуры экологических сообществ.

Ход работы:

1. Подготовка данных: матрица «виды × площадки», переменные среды.
2. Индексы альфа-разнообразия (видовое богатство, Шеннон, Симпсон).
3. Ординация PCA (rda, decostand). Визуализация, интерпретация.
4. NMDS (metaMDS). Оценка стресса. Визуализация с векторами средовых переменных (envfit).
5. Кластерный анализ (hclust, метод Уорда). Дендрограмма, выделение групп.
6. Сравнение результатов. Выводы о структуре сообщества.

Практическая работа № 5. 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.

Ход работы:

1. Загрузка генетических данных (adegenet, genind).
2. Расчёт частот аллелей и гетерозиготности (H_o , H_e).
3. Тест Харди–Вайнберга (hwe.test). Интерпретация отклонений.
4. Расчёт F-статистик (F_{is} , F_{it} , F_{st}) через hierfstat.
5. Парные F_{st} . Тепловая карта.
6. Визуализация генетической структуры (PCA, DAPC).
7. Выводы об уровне дифференциации и генетического разнообразия.

3.5. Вопросы для опроса и собеседования

1. Что показывает p-уровень значимости в статистическом тесте? Как его интерпретируют в контексте нулевой гипотезы?
2. Какие меры центральной тенденции и разброса вы знаете? В каких ситуациях предпочтительнее использовать медиану вместо среднего?
3. Какие программные пакеты в R используются для работы с генетическими данными?
4. Как рассчитать ожидаемые частоты генотипов по наблюдаемым частотам аллелей? Что означает отклонение от равновесия Харди–Вайнберга?
5. Что такое стресс в NMDS? Какие значения считаются приемлемыми, и как можно снизить стресс?
6. В чём суть метода главных компонент (PCA)? Какая информация содержится в собственных значениях и собственных векторах?

7. Чем отличается неметрическое многомерное шкалирование (NMDS) от PCA? В каких случаях NMDS предпочтительнее?
8. Что показывает коэффициент детерминации R^2 ? В чём его ограничения и почему для сравнения моделей часто используют скорректированный R^2 ?
9. Какие основные допущения лежат в основе линейной регрессионной модели? Как диагностировать их нарушение в R (графически и тестами)?
10. Что такое функция связи в обобщённых линейных моделях? Приведите примеры для нормального, биномиального и пуассоновского распределений.
11. Как интерпретируется коэффициент в логистической регрессии в терминах отношения шансов?
12. Что такое сверххрасеяние в пуассоновской регрессии и как с ним бороться?
13. Как оценить качество бинарной классификационной модели с помощью ROC-кривой? Что характеризует показатель AUC?
14. Какие индексы альфа-разнообразия вы знаете? Как они учитывают одновременно богатство и выравненность видов?
15. В чём суть метода k-средних и как выбрать число кластеров?
16. Что такое информационные критерии AIC и BIC? Для чего они используются?
17. Опишите логику проверки статистических гипотез. Что такое ошибки I и II рода?
18. В чём отличие параметрических тестов от непараметрических? Приведите примеры.
19. Что такое мультиколлинеарность и как её обнаружить в R?
20. Как интерпретировать взаимодействие факторов в двухфакторном ANOVA?
21. Что такое избыточный анализ (RDA) и как он используется в экологии?
22. Какие меры генетической дифференциации существуют помимо F_{st} ?
23. Что такое PERMANOVA и для чего она применяется?
24. Каким образом в R можно создать воспроизводимый отчёт?
25. Что такое эффективный размер популяции (N_e) и как его оценивают?
26. Какие функции в R используются для импорта данных из Excel?
27. Как в R проверить, что данные распределены нормально?
28. Что такое финальный размер эпидемии в модели SIR и как его оценить статистически?
29. Как в R можно визуализировать генетическую структуру популяций?
30. Что такое AMOVA и какие задачи она решает?
31. Какие методы регуляризации используются в регрессионных моделях (LASSO, ridge)?
32. Как в R работать с пропущенными данными (NA)?
33. Что такое бутстрап и для чего он используется в статистике?
34. Какие типы переменных различают в статистическом анализе?
35. Как интерпретировать доверительный интервал для среднего?
36. Какие пакеты в R используются для смешанных линейных моделей?
37. В чём отличие между моделями с фиксированными и случайными эффектами?
38. Как в R строятся тепловые карты и для каких данных они применяются?
39. Что такое индекс Симпсона и что он измеряет?
40. Как проверить наличие автокорреляции в данных временного ряда?
41. Как в R можно выявить и обработать выбросы?
42. Что такое байесовский подход в статистике и какие пакеты его реализуют?
43. Как интерпретировать значение F-статистики в ANOVA?
44. Какие расстояния обычно используются в экологических исследованиях (Bray–Curtis, Jaccard)?
45. Как в R провести поправку на множественные сравнения?
46. Что такое PCA по частотам аллелей и как интерпретировать результаты?
47. Как в R работать с датафреймами (пакет dplyr)?
48. Какие методы оценки параметров модели существуют (ML, REML)?
49. Как в R проверить независимость наблюдений?

50. Что такое «эффект основателя» и как он проявляется в генетических данных?

3.6. Вопросы 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. Сверхрассеяние 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. Невметрическое 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 genprop.
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).

4. Методические материалы, определяющие процедуры оценивания

4.1. Общие положения

Текущий контроль и промежуточная аттестация по дисциплине «Основы математической статистики эколого-генетических систем на языке программирования R» проводятся в соответствии с балльно-рейтинговой системой Университета. Цель оценочных процедур – проверка уровня сформированности компетенций **СПК-2, СПК-4** на различных этапах обучения.

4.2. Формы текущего контроля и их балльная оценка

Форма контроля	Количество за семестр	Максимальный балл за единицу	Максимальный балл за семестр
Опрос/собеседование	2	5	10
Тестирование	2	10	20

Форма контроля	Количество за семестр	Максимальный балл за единицу	Максимальный балл за семестр
Практическая работа	5	10	50
Итого за текущий контроль			80

4.3. Процедура проведения зачёта

Зачёт проводится в период сессии в устной форме. Билет содержит два теоретических вопроса и одно практическое задание (написание фрагмента кода на R или интерпретация результатов). Время на подготовку – 30 минут. Ответ должен быть полным, аргументированным, с использованием статистической терминологии.

Шкала оценивания зачёта

Показатель	Баллы
Отлично (высокий уровень). Студент демонстрирует всестороннее, глубокое и систематическое знание теоретических основ статистического анализа, свободно ориентируется в современных пакетах R. Ответы полные, логически безупречные, подкреплены формулами и примерами. Практическое задание выполнено безошибочно. Студент способен объяснить допущения модели, предложить альтернативные подходы.	17–20
Хорошо (базовый уровень). Студент показывает хорошее знание материала, основные теоретические положения излагает верно. Ответы достаточно полны, но могут содержать незначительные неточности. Практическое задание выполнено в целом правильно, но возможны небольшие погрешности. Отвечает на большинство дополнительных вопросов.	13–16
Удовлетворительно (пороговый уровень). Студент владеет материалом на базовом уровне, но знания фрагментарны. Ответы кратки, требуют наводящих вопросов, содержат отдельные ошибки. Практическое задание выполнено с ошибками: выбранный метод не вполне адекватен данным, интерпретация поверхностна. Студент может объяснить лишь основные этапы анализа.	9–12
Неудовлетворительно (низкий уровень). Студент демонстрирует существенные пробелы в знаниях. Теоретические ответы отрывочны, содержат грубые ошибки, основные понятия искажены. Практическое задание либо не выполнено, либо содержит принципиальные ошибки. Студент не может ответить на дополнительные вопросы.	5–8
Крайне неудовлетворительно. Не отвечает на теоретические вопросы или ответы полностью не соответствуют теме. Практическое задание не выполнено. Не владеет даже минимальными сведениями.	0–4

Максимальное количество баллов на зачёте – 20.

Итоговая шкала оценивания по дисциплине

Сумма баллов	Оценка
81–100	зачтено
61–80	зачтено
41–60	зачтено
0–40	не зачтено

4.4. Порядок ликвидации академической задолженности

Студенты, набравшие менее 11 баллов текущего контроля или не сдавшие зачёт, проходят пересдачу в порядке, установленном локальными актами Университета. Пересдача включает выполнение дополнительных заданий (повторное тестирование, выполнение пропущенных практических работ, собеседование).

Ключи ко всем заданиям приведены непосредственно в тексте (для тестов, ситуационных задач, открытых вопросов). Для теоретических вопросов (опрос, зачёт) критерии оценивания указаны в разделе 2; преподаватель оценивает полноту и правильность ответа.