

**Федеральное государственное бюджетное учреждение науки
Институт проблем экологии и эволюции им. А.Н. Северцова Российской
академии наук**



Утверждаю.
Директор ИПЭЭ РАН


Найденко С.В.
«30» апреля 2026 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Статистический анализ данных в биологии»

Группа специальностей:

1.5 Биологические науки

Специальности:

«Зоология», «Энтомология», «Ихтиология», «Экология», «Гидробиология», «Паразитология»

Москва, 2026 г.

Аннотация

Дисциплина «**Статистический анализ данных в биологии**» реализуется в рамках основной профессиональной образовательной программы высшего образования - программы подготовки научно-педагогических кадров в аспирантуре Федерального государственного бюджетного учреждения науки Института проблем экологии и эволюции им. А.Н. Северцова Российской академии наук (ИПЭЭ РАН) по группе специальностей «Биологические науки».

Основным источником материалов для формирования содержания программы являются: учебники, монографические издания, публикации, материалы конференций, симпозиумов, семинаров, интернет-ресурсы. Общая трудоемкость дисциплины по учебному плану составляет 3 зачетных единицы, 108 часов работы. Дисциплина реализуется на 2 году обучения. Текущая аттестация проводится не менее 2 раз в соответствии с заданиями и формами контроля, предусмотренные настоящей программой. Промежуточная оценка знания осуществляется в форме экзамена.

ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ: овладение теоретическими основами и инструментарием современного математического анализа данных, в частности, в программной среде для обработки данных R (R Development Core Team).

ЗАДАЧИ ДИСЦИПЛИНЫ.

В результате освоения дисциплины аспирант должен:

Знать:

требования к выборкам и ограничения, которые задаются предполагаемыми методами анализа данных, основные теоретические принципы анализа данных, тестирования статистических гипотез, принципы построения аналитических моделей, подбора оптимальных моделей

Уметь:

подготавливать данные для последующего анализа, формировать выборки под конкретные исследовательские задачи, производить анализ данных с использованием специализированного программного обеспечения.

Владеть:

представлениями о статистических методах и критериях, используемых в современных биологических исследованиях, об их назначении, ограничениях и об интерпретации результатов статистических тестов; современными методами статистического анализа данных.

ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов.

СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

№ п/п		Наименование раздела дисциплины	Общая трудоемкост в академ. часах
1	Основные понятия в анализе биологических данных. Описательная статистика. Тестирование гипотез	Статистика как инструмент для количественного анализа и интерпретации биологических данных. Понятия выборки и измерения. Принципы сбора данных для статистического анализа. Случайность и независимость измерений. Частотное распределение переменной. Характеристики распределения (центр, разброс и форма распределения). Нормальное распределение. Распределение выборочных средних. Понятие статистики критерия. Понятие статистической гипотезы, принципы формулирования взаимоисключающих гипотез. Процедура тестирования гипотез в статистике, статистические ошибки. Критерии Стьюдента. Проверка соответствия распределения в выборке нормальному закону. Гомогенность дисперсии. Тесты на гомогенность. Выбросы (outliers). Биномиальное распределение. Распределение Пуассона. Непараметрические критерии.	18
2	Основы анализа данных с помощью языка R.	Язык R как аппарат для анализа биологических данных. Подготовка данных для анализа в R и импорт данных. Графические интерфейсы в R. Основы синтаксиса. Подбор и загрузка программных пакетов и поиск функций, соответствующих задачам исследования. Описание данных. Основные статистические критерии в R. Преимущества R в построении сложных аналитических моделей, обзор возможностей R, выходящих за пределы возможностей традиционного программного обеспечения.	12
3	Дисперсионный анализ (ANOVA).	Дисперсионный анализ. Рекомендации и требования к выборкам. Однофакторный дисперсионный анализ. Эффект множественных сравнений. Апостериорные тесты. Анализ контрастов. Дисперсионный анализ для связанных выборок. Многофакторный дисперсионный анализ.	10

№ п/п		Наименование раздела дисциплины	Общая трудоемкост в академ. часах
4	Корреляции. Регрессионный анализ.	Коэффициент корреляции как параметр описательной статистики. Коэффициент корреляции Пирсона. Определение знака и величины коэффициента корреляции. Факторы, влияющие на коэффициент корреляции. Получение матрицы корреляции. Регрессионный анализ. Уравнение регрессии. Ошибки предсказания и поиск лучшей линии регрессии. Метод наименьших квадратов. «Анализ остатков» (residual analysis). Множественная линейная регрессия и корреляция.	10
5	Обобщённые линейные модели.	Обобщённые линейные модели (GLM) и их применение в биологических и экологических исследованиях. Принципы построения модели. Тестирование гипотез. Поиск оптимальной модели и информационные критерии. Ранжирование моделей с помощью информационного критерия Акаике. Принципы работы с программным пакетом MuMIN в среде R. Обобщенные линейные модели для счетных данных. Оценка избыточной дисперсии. Обобщенные линейные модели с бинарным откликом. Моделирование вероятностей и долей.	18
6	Обобщенные линейные модели со смешанными эффектами	Обобщенные линейные модели со смешанными эффектами (GLMM) для иерархически организованных данных. GLMM для нормально-распределенных данных. Программные пакеты lme4 и nlme в среде R. Моделирование структуры дисперсии в смешанных моделях. GLMM для счетных данных. GLMM для бинарных данных.	12
7	Факторный анализ. Анализ главных компонент.	Анализ главных компонент, его основные этапы. Главные компоненты, их структура и интерпретация. Вращение компонент. Использование метода главных компонент для решения проблемы коллинеарности переменных в линейных моделях. Реализация метода в среде R. Применение метода главных компонент в биологии.	12
8	Дискриминантный анализ.	Дискриминантные функции, их структура и интерпретация. Функции классификации и Классификационная матрица. Пошаговый дискриминантный анализ. Обобщённый дискриминантный анализ. Применение дискриминантного анализа в экологических исследованиях. Реализация метода в среде R.	12

№ п/п		Наименование раздела дисциплины	Общая трудоемкост в академ. часах
9		Экзамен	4
		ИТОГО	108

Образовательные технологии

Лекции, семинары, практические занятия, написание рефератов, подготовка презентаций и выступлений.

Текущая и промежуточная аттестация

Текущая аттестация по дисциплине проводится в форме решения задач по данной дисциплине. Объектами оценивания выступают: активность на занятиях, своевременность выполнения различных видов заданий, посещаемость занятий; степень усвоения теоретических знаний и уровень овладения практическими умениями и навыками по всем видам учебной работы, проводимых в рамках семинаров, практических занятий и самостоятельной работы.

Промежуточная аттестация по дисциплине осуществляется в форме экзамена. Аспирант допускается к экзамену в случае выполнения всех учебных заданий и мероприятий, предусмотренных настоящей программой. В случае наличия учебной задолженности (пропущенных занятий и (или) невыполненных заданий) аспирант отрабатывает пропущенные занятия и выполняет задания.

Оценивание обучающегося на промежуточной аттестации осуществляется на экзамене с использованием нормативных оценок по 5-х бальной системе (5 - отлично, 4 - хорошо, 3 - удовлетворительно, 2 – неудовлетворительно).

Экзамен проводится по билетам, в которых предлагается 2 вопроса.

Оценивание аспиранта на промежуточной аттестации в форме экзамена

Оценка	Требования к знаниям и критерии выставления оценок
2, неудовлетворительно	Аспирант при ответе демонстрирует плохое знание значительной части основного материала в области статистического анализа данных. Не информирован или слабо разбирается в проблемах и/или не в состоянии наметить пути их решения, не может статистически проанализировать данные в соответствии с поставленной задачей.
3, удовлетворительно	Аспирант при ответе демонстрирует знания только основного материала в области статистического анализа данных, фрагментарно понимает назначение основных статистических методов, их ограничения и интерпретацию. Не всегда может подобрать корректный метод анализа данных для заданной ситуации, провести соответствующий анализ с использованием специализированного программного обеспечения, интерпретировать результаты и изложить их в соответствии с общепринятыми стандартами
4, хорошо	Поступающий при ответе демонстрирует хорошие знания в области статистического анализа данных, владеет основными принципами статистического анализа данных, понимает назначение основных статистических методов, их ограничения и интерпретацию. Может подобрать корректный метод анализа данных для заданной ситуации, провести соответствующий анализ с использованием специализированного программного обеспечения, но не всегда в состоянии интерпретировать результаты и изложить их в соответствии с общепринятыми стандартами.

5, отлично	Поступающий при ответе демонстрирует глубокое и прочное владение и использование знаний в области статистического анализа данных, владеет принципами статистического анализа данных, понимает назначение статистических методов, их ограничения и интерпретацию. Информирован и способен делать анализ проблем и намечать пути их решения. Может подобрать корректный метод анализа данных для заданной ситуации, провести соответствующий анализ с использованием специализированного программного обеспечения, интерпретировать результаты и изложить их в соответствии с общепринятыми стандартами.
------------	--

Экзаменационные вопросы

1. Статистика как инструмент для количественного анализа и интерпретации данных. Описательная и индуктивная статистика.
2. Понятия выборки и измерения. Принципы сбора данных для статистического анализа. Свойства переменных. Частотное распределение переменной.
3. Графическое представление частотного распределения.
4. Характеристики распределения.
5. Нормальное распределение. Площадь нормального распределения.
6. Распределение выборочных средних.
7. Понятие статистики критерия. Понятие статистической гипотезы, принципы формулирования взаимоисключающих гипотез.
8. Процедура тестирования гипотез в статистике, статистические ошибки.
9. Критерии Стьюдента: одновыборочный, двухвыборочный для независимых выборок, двухвыборочный для связанных выборок.
10. Дисперсионный анализ. Связь с двухвыборочным t-критерием Стьюдента. Рекомендации и требования к выборкам.
11. Однофакторный дисперсионный анализ.
12. Эффект множественных сравнений. Апостериорные тесты (Бонферрони, Тьюки, Ньюмена-Кейлса, Даннета, Шеффе).
13. Анализ контрастов.
14. Дисперсионный анализ для связанных выборок.
15. Многофакторный дисперсионный анализ.
16. Модели дисперсионного анализа. Модель смешанных эффектов.
17. Гнездовой дисперсионный анализ.
18. Многомерный дисперсионный анализ.
19. Понятие мощности статистического критерия.
20. Практическая интерпретация результатов (размер эффекта).
21. Оценка необходимого размера выборки для предполагаемого статистического анализа.
22. Коэффициент корреляции как параметр описательной статистики. Коэффициент корреляции Пирсона.
23. Определение знака и величины коэффициента корреляции. Факторы, влияющие на коэффициент корреляции.
24. Получение матрицы корреляции. Требование к выборке для тестирования гипотезы о коэффициенте корреляции Пирсона.
25. Регрессионный анализ. Уравнение регрессии.
26. Ошибки предсказания и поиск лучшей линии регрессии. Метод наименьших квадратов.
27. Анализ остатков.
28. Множественная линейная регрессия и корреляция.
29. Нелинейная регрессия.
30. Анализ ковариаций.
31. Требования к выборке для проведения параметрических тестов. Случайность измерений. Независимость измерений.

32. Проверка соответствия распределения в выборке нормальному закону. Критерии Колмогорова-Смирнова, Лиллифорса, Шапиро-Уилкса.
33. Гомогенность дисперсии. Тесты на гомогенность.
34. Выбросы (outliers).
35. Биномиальное распределение. Распределение Пуассона.
36. Трансформация данных.
37. Непараметрические критерии. Ранжирование данных.
38. Критерии Манна-Уитни, Вальда-Вольфовица, Вилкоксона, Крускала-Уоллеса, Фридмана, знаковый, медианный тесты.
39. Ранговые корреляции.
40. Критерии согласия. Сравнение наблюдаемых частот с теоретическими.
41. Биномиальный тест.
42. Критерий χ^2 Пирсона. Поправка Йейтса.
43. Таблицы сопряжённости. Четырёхпольные таблицы для независимых выборок.
44. Точный критерий Фишера и критерий χ^2 для независимых выборок.
45. Корреляция между качественными переменными.
46. Повторные измерения бинарных переменных: критерии Мак-Немара и Кохрана.
47. Лог-линейные модели.
48. Общие представления о методах многомерного анализа. Понятие многомерного распределения, его характеристики.
49. Матрицы дисперсий, ковариаций и корреляций переменных. Собственное значение и собственный вектор матрицы. Матрица дистанций между объектами.
50. Подготовка данных для многомерного анализа.
51. Дискриминантный анализ. Дискриминантные функции, их структура и интерпретация.
52. Функции классификации и Классификационная матрица.
53. Пошаговый дискриминантный анализ.
54. Обобщённый дискриминантный анализ.
55. Задачи факторного анализа.
56. Различные типы факторного анализа. Подготовка данных к анализу.
57. Анализ главных компонент, его основные этапы.
58. Главные компоненты, их структура и интерпретация.
59. Вращение компонент.
60. Канонический анализ.
61. Выявление структуры данных на основе матриц дистанций между объектами. Способы оценки дистанций в многомерном пространстве.
62. Многомерное шкалирование, основные этапы.
63. Диаграмма Шепарда. Интерпретация результатов многомерного шкалирования.
64. Кластерный анализ. Различные методики кластерного анализа, варианты построения деревьев классификации.
65. Кластеризация методом К средних.
66. Обобщённые линейные модели. Принципы построения модели.
67. Поиск оптимальной модели и информационные критерии.
68. Язык R как аппарат для анализа биологических данных. Подготовка данных для анализа в R и импорт данных.
69. Графические интерфейсы в R.
70. Основы синтаксиса в R.
71. Подбор и загрузка программных пакетов и поиск функций, соответствующих задачам исследования.
72. Описание данных. Основные статистические критерии в R.

Учебно-методическое и информационное обеспечение дисциплины:
Основная литература

1. Sokal, R. R. and F. J. Rohlf. 2012. Biometry: the principles and practice of statistics in biological research. 4th edition. W. H. Freeman and Co.: New York. 937 pp.
2. Zar J.K., 2010. Biostatistical analysis. 5th ed. (или 4d ed, 1999) Prentice Hall, New Jersey. 944 P.
3. Zuur, A. F. Mixed Effects Models and Extensions in Ecology with R / A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, G. M. Smith. — New York : Springer, 2009. — 574 p. — ISBN 978-0-387-87457-9.
4. Гланц С, 1998. Медико-Биологическая Статистика McGraw-Hill, 1994; М.: Практика. 459 с.
5. Ивантер, Э. В. Введение в количественную биологию : учебное пособие для студентов биологических специальностей / Э. В. Ивантер, А. В. Коросов ; Петрозавод. гос. ун-т. — 3-е изд., испр. и доп. — Петрозаводск : Изд-во ПетрГУ, 2014. — 298 с. : табл., рис. — ISBN 978-5-8021-2134
6. Кабаков, Р. И. R в действии : анализ и визуализация данных с использованием R и Tidyverse : 3-е изд. / Р. И. Кабаков — Москва : ДМК Пресс, 2023. — 768 с. — ISBN 978-5-93700-173-3.
7. Лакин Г. Ф. Биометрия: учебное пособие для биол. спец. вузов //М.: Высшая школа. — 1990. — Т. 352.
8. Мاستицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. - М.: ДМК Пресс, 2015. - 496 с. <http://www.ievbras.ru/ecostat/Kiril/>

Дополнительная литература

1. Hurlburt R.T., 2006. Comprehending behavioral statistics. 4rd ed. (или 3rd ed, 2003), Wadsworth/Thomson Learning, Belmont.
2. Lehner P.N., 1996. Handbook of ethological methods. Cambridge University press. 672p.
3. Боровиков В. Популярное введение в программу Statistica. КомпьютерПресс. 2000.
4. Тихомирова М. М. (ред.). Биометрия: учебное пособие. — Изд-во Ленингр. ун-та, 1982.

Базовые журналы:

1. Биометрика
2. Доклады РАН
3. Известия РАН
4. Статистические методы
5. Успехи современной биологии

Библиотечные и Интернет-ресурсы

№	Ссылка на информационный ресурс	Наименование разработки в электронной форме	Доступность (количество точек доступа)
1	http://www.nature.com/nature	Nature	64
2	http://www.nature.com/methods	Nature Methods	64
3	http://www.webofknowledge.com	Web of Science. Библиографическая база данных	64
4	http://www.sciencedirect.com/ science	ScienceDirect. База журналов издательства Elsevier	64
5	http://www.elsevier.com	Elsevier Поисковая система публикаций	64
6	http://www.springerlink.com	SpringerLink. База журналов издательства Springer	64

7	http://www.springer.com	Springer Поисковая система публикаций	64
8	http://www.annualreviewws.org	Annual Reviews. База	64
9	http://onlinelibrary.wiley.com/	Wiley Электронная библиотека	64
10	http://online.sagepub.com/	Sage Journals	64
11	http://www.annualreviews.org/	Annual Reviews Sciences Collection	64
12	http://www.sciencemag.org/journals	Science/AAAS	64

Материально-техническое обеспечение дисциплины.

В ИПЭЭ РАН имеется следующее оборудование: компьютеры со специализированным программным обеспечением; проектор, сеть WiFi, ноутбуки, плазменные панели и проекторы для демонстраций

Учебно-методическое обеспечение самостоятельной работы.

Библиотечные и Интернет-ресурсы, консультации с ведущими специалистами Института.

Язык преподавания: русский.

Авторы программы: к.б.н. Васильева Н.А., к.б.н. Шитиков Д.А.

Программа одобрена на заседании Ученого совета ИПЭЭ РАН, протокол № _____ от _____ .2026 года.