Приложение № 4
к образовательной программе высшего образования – программе магистратуры
по направлению подготовки 06.04.01 Биология
(направленность (профиль) «Генетика и генетические технологии»)

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ «НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»
(АНОО ВО «УНИВЕРСИТЕТ «СИРИУС»)

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат: 7835bcdac1c89a3da835a18f6045494f3f404337
Владелец: Некрасова Людмила Андреевна
Действителен: с 08.07.2026 до 08.07.2027

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Биостатистика с применением языка R»

1. Трудоемкость дисциплины (модуля): 3 з.е.

2. Место дисциплины (модуля) в учебном плане:

Обязательная часть «Профессиональной подготовки» в рамках Блока 1. Дисциплины (модули), реализуется во 2 семестре.

3. Цель дисциплины (модуля): формирование теоретических знаний и практических навыков в области статистической обработки экспериментальных данных, полученных в ходе проведения биологических экспериментов, включая молекулярно-генетические исследования, на языке программирования R.

4. Задачи дисциплины (модуля):

– научить выбирать метод статистической обработки данных, соответствующий поставленной задаче;

– сформировать представление об основных методах статистической обработки данных на языке программирования R;

– научить оценивать достоверность выводов, полученных в ходе анализа экспериментальных данных;

– научить визуализировать исходные данные и результаты статистической обработки данных в графической среде R.

5. Перечень разделов (тем) дисциплины (модуля) и их краткое содержание:

Наименование раздела (темы) дисциплины (модуля)	**Краткое содержание**
Раздел 1. Введение	Работа с исходными данными с использованием R. Полный цикл предварительной обработки данных: импорт из внешних источников (CSV, Excel, SPSS, базы данных через DBI, RSQLite), первичный просмотр структуры (str(), glimpse()), приведение типов, создание факторов, удаление дубликатов и обработку пропущенных значений (na.omit, tidyr::replace_na, mice для множественного восполнения).
Раздел 2. Статистические тесты.	Статистические тесты, гипотезы (H0, H1), ошибки (альфа, бета). Типы ошибок: ошибка I рода (α) — ложное отклонение H_0 (вероятность ложноположительного результата), ошибка II рода (β) — ложное принятие H_0, мощность теста = 1-β.

Раздел 3. Анализ многомерных таблиц.	Анализ многомерных таблиц. Таблицы сопряженности. Построение в R. Критерий χ^2 Пирсона, точный критерий Фишера. Для многомерных таблиц (три и более переменных) – изучение частичных таблиц, использование логарифмически-линейных моделей .
Раздел 4. Регрессионный анализ и ANOVA.	Линейная регрессия (lm): спецификация модели, оценка коэффициентов (МНК), интерпретация коэффициентов (влияние предиктора на отклик), summary(), доверительные интервалы для коэффициентов (confint). Диагностика модели: график остатков, QQ-plot, проверка гетероскедастичности, мультиколлинеарности. ANOVA (однофакторный, двухфакторный, с повторными измерениями).
Раздел 5. Непараметрические тесты.	Поправка на множественную проверку гипотез. Многомерная статистика. Визуализация многомерных данных без сокращения размерности.
Раздел 6. Многомерное шкалирование.	Базовые методы сокращения размерности (PCA, MDS). Реализация в R (с центрированием и масштабированием).

6. Образовательные результаты освоения дисциплины (модуля):

Код и наименование компетенции	**Код и наименование индикатора достижения компетенции**
ПК-2. Способен применять методы молекулярной диагностики наследственных заболеваний и использовать молекулярно-генетические методы исследования в области медицины и биотехнологии	ИПК-2.3. Собирает, отбирает и использует необходимые данные и эффективно применяет количественные методы их анализа
ПК-5. Способен использовать современные информационные технологии, в том числе специализированное программное обеспечение, для решения прикладных задач в научно-исследовательской (проектно-исследовательской) деятельности	ИПК-5.2. Умеет самостоятельно осуществлять обработку и анализ геномных данных с использованием современных информационных технологий, в том числе языков программирования R и Python
	ИПК-5.3. Использует современные информационные технологии, в том числе специализированное программное обеспечение, для проведения биоинформатического анализа больших

	объемов геномных данных