

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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Теория вероятностей и математическая статистика»

Уровень образования:	высшее образование – программа специалитета
Специальность:	06.05.01 Биоинженерия и биоинформатика
Направленность (профиль):	Биоинженерия

1. **Трудоемкость дисциплины (модуля):** 4 з.е.

2. **Место дисциплины в учебном плане:** дисциплина «Теория вероятностей и математическая статистика» входит в Блок 1. «Дисциплины (модули)», обязательную часть, раздел «Профессиональная подготовка» и изучается в 7-10 модулях (3-4 семестры).

3. **Цель дисциплины (модуля):** сформировать у студентов представления о закономерностях случайных явлений и научить применять вероятностно-статистические методы для анализа и интерпретации данных для решения прикладных задач в биоинженерии и смежных науках.

4. **Задачи дисциплины (модуля):**

- Изучение основных понятий теории вероятностей (события, вероятности, условная вероятность).
- Понимание стохастических процессов в биологии: моделирование случайных процессов в биологических системах (экспрессия генов, мутации, эпидемиологические процессы).
- Освоение дискретных и непрерывных распределений, применяемых в биологии.

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

Дисциплина знакомит студентов с базовыми понятиями теории вероятностей и математической статистики. Курс построен на прикладных примерах из биоинженерии, биологии и биоинформатики и включает освоение инструментов визуализации и анализа в языке программирования R. Особое внимание уделяется интерпретации результатов, корректной постановке задач и пониманию ограничений статистических методов. Освоение курса способствует формированию у обучающихся системного статистического мышления, необходимого для анализа данных в биомедицинских исследованиях.

Раздел	Содержание
Раздел 1. Введение. Данные и инструменты анализа	Типы биологических данных. Природа случайности в экспериментах. Структура и цели курса. Формализация данных и переменных. Основы синтаксиса языка R, описание векторов, факторов, датафреймов.
Раздел 2. Основы теории вероятностей	Аксиомы вероятности. Пространство элементарных событий. События и операции над ними. Классическое определение вероятности. Комбинаторика. Условная вероятность, формула Байеса. Независимость. Схема Бернулли. Моделирование случайных процессов с помощью R.
Раздел 3. Случайные величины и распределения	Случайные величины, функции распределения, плотности и квантили. Дискретные и непрерывные распределения: биномиальное, нормальное, пуассоновское, экспоненциальное. Математическое ожидание, дисперсия. Закон больших чисел. Центральная предельная теорема. Симуляции распределений, наложение плотностей, генерация данных в R.
Раздел 4. Первичный анализ и визуализация данных	Выборка и её характеристики. Вариационные ряды, гистограммы, boxplot. Характеристики распределения: среднее, медиана, мода, стандартное отклонение. Эмпирическая функция распределения. Проверка на нормальность (Q-Q plot, Shapiro-Wilk). Ковариация и коэффициент корреляции. Работа с биологическими данными в R, визуализация и предварительный анализ.
Раздел 5. Оценка параметров и проверка гипотез	Точечные оценки: метод моментов, метод максимального правдоподобия. Интервальные оценки. Проверка статистических гипотез: нулевая и альтернативная гипотезы, ошибки I и II рода, p-value. Т-критерий, непараметрические тесты (тест Вилкоксона и Манна-Уитни), критерий хи-квадрат, критерий Колмогорова–Смирнова. Поправки на множественное сравнение.
Раздел 6. Регрессионный	Простая и множественная линейная регрессия. Дисперсионный анализ (ANOVA). Анализ ковариации (ANCOVA). Обобщённые линейные модели: логистическая регрессия и модели для счётных данных. Анализ

анализ и прикладные модели		качества модели. Построение моделей, интерпретация результатов, работа с реальными данными из биоинженерии в R.
Раздел	7.	Планирование эксперимента: рандомизация, блок-дизайн, контроль батч-эффектов, расчёт мощности и выбор оптимального размера выборки.
Планирование эксперимента		

6. Образовательные результаты освоения дисциплины (модуля):

Формируемые компетенции (код компетенции, формулировка)	Планируемые результаты обучения по дисциплине (индикаторы достижения компетенций)
ОПК-2. Способен использовать специализированные знания фундаментальных разделов математики, физики, химии и биологии для проведения исследований в области биоинженерии, биоинформатики и смежных дисциплин (модулей)	ИОПК-2.1 Применяет вероятностно-статистические методы к анализу биологических и биоинженерных данных
	ИОПК-2.2 Интерпретирует результаты статистических тестов и моделей, выбирает адекватные методы анализа с учётом типа данных и исследуемой задачи

7. Оценочные и методические материалы

7.1. Оценочные материалы для организации текущего контроля

Контрольные работы (КР1 – КР2)

Форма: письменная, синхронная

Место и время проведения: во время контактной работы в аудитории, согласно расписанию

Примеры заданий (по парам):

КР1 (3 семестр). Основы теории вероятностей

I. Вероятность событий

1. В эксперименте по секвенированию ДНК вероятность ошибки чтения нуклеотида равна 0.001. Какова вероятность, что в цепочке из 1000 нуклеотидов не будет ни одной ошибки?

2. В популяции 5% людей имеют генетическую мутацию. Какова вероятность, что среди 20 случайно выбранных человек окажется ровно 2 носителя?

II. Условная вероятность

1. Тест на заболевание дает 2% ложноположительных и 1% ложноотрицательных результатов. Если 0.5% популяции больны, какова вероятность, что пациент с положительным результатом действительно болен?

2. Вероятность экспрессии гена А — 0.6, гена В — 0.4. Если известно, что ген В экспрессируется, какова вероятность одновременной экспрессии А и В?

III. Случайные величины

1. Число мутаций в геноме описывается распределением Пуассона со средним $\lambda=3$. Найдите вероятность ≥ 2 мутаций.

2. Длина белковой цепи распределена нормально с $\mu=100$ а.о. и $\sigma=5$. Какая доля белков имеет длину от 90 до 110 а.о.?

КР2 (4 семестр). Математическая статистика

I. Описательная статистика

1. Для данных экспрессии гена (в усл. ед.): [12, 15, 18, 14, 16], вычислите:

- Выборочное среднее и дисперсию
- Медиану и интерквартильный размах

2. Проверка гипотез

Сравните две группы пациентов по уровню белка (мг/мл):

Группа 1: [45, 50, 55, 48]

Группа 2: [52, 58, 60, 54]

Проверьте гипотезу о равенстве средних ($\alpha=0.05$).

В эксперименте с лекарством 18 из 25 пациентов выздоровели. Можно ли утверждать, что эффективность $>60\%$ ($\alpha=0.05$)?

3. Корреляция и регрессия

Для данных:

Доза препарата (X)	1	2	3	4
Эффект (Y)	5	7	9	11
Постройте линейную регрессию Y на X.				

Критерии оценки:

1. Корректность выполнения заданий — до 4 баллов.

2. Полнота и логика — до 4 баллов.

Доклады-презентации (Д1 – Д4)

Форма: устная, синхронная

Место и время проведения: очно, в аудитории, в рамках практического занятия

Примеры тем:

1. Вероятностные модели в биологии: Стохастические процессы в экспрессии генов.

Моделирование случайных флуктуаций mRNA и белков.

2. Вероятность ошибок при секвенировании ДНК. Распределение Пуассона для анализа ошибок в NGS-данных.

3. Нейросети в обработке биомедицинских изображений: распознавание патологий на МРТ.

4. Статистическая оценка качества сборки генома: N50, контиги и другие метрики.

5. Применение статистики в медицинских исследованиях: анализ выживаемости (Survival Analysis) в клинических испытаниях.

6. Логистическая регрессия для предсказания заболеваний на примере риска диабета по генетическим маркерам.

Критерии оценки:

7. Структура и логика доклада — до 4 баллов

8. Языковая грамотность — до 3 баллов

9. Ответы на вопросы, аргументация — до 3 баллов

Пересдача: Не более двух раз

Проект (П)

Форма: очная защита проекта

Место проведения: в аудитории, в рамках итогового занятия

Время проведения: 8 модуль / 4 семестр

Примеры групповых заданий:

1. Статистическая обработка данных RNA-seq

Задача: Сравнить экспрессию генов между двумя группами образцов (например, здоровые vs больные).

Методы:

Нормализация данных (FPKM, TPM).

Дифференциальная экспрессия (DESeq2, edgeR).

Поправка на множественное тестирование (FDR).

Выход: Таблица значимых генов, volcano plot.

2. Анализ частот аллелей в популяции

Задача: Проверить равновесие Харди-Вайнберга для SNP (данные из 1000 Genomes Project).

Методы:

Критерий χ^2 .

Визуализация распределения генотипов.

Выход: Отчет с p-value и интерпретацией.

3. Моделирование эпидемии

Задача: Спрогнозировать распространение инфекции с помощью SIR-модели.

Параметры: R_0 , время выздоровления.

Выход: Графики динамики зараженных.

Критерии оценки:

Содержание и научная значимость — до 4 баллов

Логика и структура изложения — до 3 баллов

Владение терминологией, ответы на вопросы — до 3 баллов

7.2. Оценочные материалы для организации промежуточной аттестации

- Форма проведения: письменная (синхронная), в очном формате в зависимости от расписания. Проведение осуществляется в рамках зачетной недели во втором и четвертом семестре (4 и 8 модуль дисциплины соответственно).

- Место проведения: учебная аудитория

Пример экзаменационного тестового задания:

10. Если $P(A)=0.3$, $P(B)=0.4$, и события независимы, то $P(A \cup B)P(A \cup B)$ равна:

Варианты:

a) 0.7

b) 0.58

c) 0.12

d) 0.5

11. Тест на заболевание имеет:

— Чувствительность (sensitivity) = 95%

— Специфичность (specificity) = 90%

— Распространенность болезни = 1%.

Чему равна вероятность болезни при положительном результате?

Варианты:

a) ~8.7%

b) ~95%

c) ~50%

d) ~1%

3. Вероятность ошибки при секвенировании одного нуклеотида равна 0.001. Какова вероятность, что в цепочке из 100 нуклеотидов будет не более одной ошибки?

Варианты:

a) 0.9953

b) 0.9048

c) 0.9995

d) 0.3679

Критерии оценивания:

Корректность выполнения заданий — до 4 баллов;

Полнота и аргументация ответов — до 4 баллов;

Владение профессиональной лексикой — до 2 баллов.

Максимальное количество баллов за экзамен — 10.

7.3. Методические рекомендации

Обучение по дисциплине предполагает изучение курса на аудиторных занятиях (практические занятия) и в ходе самостоятельной работы студентов. Студентам необходимо ознакомиться с содержанием рабочей программы дисциплины, с целями и задачами дисциплины, ее связями с другими дисциплинами образовательной программы, методическими разработками по данной дисциплине.

Обучение по дисциплине проводится последовательно путем проведения практических занятий с углублением и закреплением полученных знаний в ходе самостоятельной работы с последующим переводом знаний в умения в ходе практических занятий. Получение углубленных знаний по изучаемой дисциплине достигается за счет дополнительных часов к аудиторной работе самостоятельной работы студентов. Выделяемые часы целесообразно использовать для знакомства с дополнительной научной литературой по проблематике дисциплины, анализа научных концепций и современных подходов к осмыслению рассматриваемых проблем. К самостоятельному виду работы студентов относится работа в библиотеках, в электронных поисковых системах и т.п. по сбору материалов, необходимых для проведения практических занятий или выполнения конкретных заданий преподавателя по изучаемым темам. Обучающиеся могут установить электронный диалог с преподавателем, выполнять посредством него контрольные задания.