

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ
УЧРЕЖДЕНИЕ «НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
БИОМЕДИЦИНСКОЙ ХИМИИ ИМЕНИ В.Н. ОРЕХОВИЧА» (ИБМХ)

ОДОБРЕНО

Решением Ученого совета

Протокол № 5 от «30» июня 2022 г.

УТВЕРЖДЕНО

Приказом директора ИБМХ

от «01» июля 2022 г. № 43

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
«СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ В БИОЛОГИИ И
МЕДИЦИНЕ»**

Научная специальность

1.5.8 Математическая биология, биоинформатика

1.5.4 Биохимия

Москва 2022 г.

Рабочая программа дисциплины (модуля) «Статистический анализ данных в биологии и медицине» разработана в соответствии с Федеральными государственными требованиями, утверждёнными приказом Министерства науки и высшего образования Российской Федерации от 20.10.2021 № 951, работниками отдела биоинформатики ИБМХ.

№	Фамилия, имя, отчество	Ученая степень, звание	Занимаемая должность в ИБМХ
1	Пятницкий Михаил Алексеевич	канд. биол. наук	в.н.с. лаб. анализа постгеномных данных
2	Филимонов Дмитрий Алексеевич	канд. ф.-м. наук	в.н.с. лаб. структурно- функционального конструирования лекарств

ОГЛАВЛЕНИЕ

1. Цель и задачи изучения дисциплины (модуля).....	4
2. Объем дисциплины (модуля) по видам учебной работы	4
3. Содержание дисциплины (модуля)	4
4. Учебно-тематический план дисциплины (модуля).....	5
5. Учебно-методическое обеспечение самостоятельной работы обучающихся.....	5
6. Оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся	6
7. Учебно-методическое и информационное обеспечение дисциплины (модуля).....	9
7. Материально-техническое обеспечение дисциплины (модуля)	9
8. Методические указания для обучающихся по изучению дисциплины (модуля)	10
9. Методические рекомендации преподавателю по организации учебного процесса по дисциплине (модулю)	10

1. Цель и задачи изучения дисциплины (модуля)

Цель изучения дисциплины (модуля)

Изучение дисциплины (модуля) «Статистический анализ данных в биологии и медицине» предоставляет получение основополагающих знаний и практических навыков в современных методах статистики, используемых для анализа данных в биологии и медицине (включая анализ омиксных данных) для подготовки научных кадров высшей квалификации по специальности 1.5.8 Математическая биология, биоинформатика и 1.5.4. Биохимия.

Задачи дисциплины (модуля)

1. Сформировать комплекс базовых знаний о статистических методах анализа биологических и медицинских данных, их областях применимости и ограничениях
2. Изучить теоретические основы базовых методов системной биологии, способы визуализации и интеграции омиксных данных.
3. Освоить практические навыки статистической обработки данных и системно-биологического анализа на языках программирования R и/или Python.

2. Объем дисциплины (модуля) по видам учебной работы

Таблица 1

Виды учебной работы		Всего, час.	Объем по семестрам							
			1	2	3	4	5	6	7	8
Контактная работа обучающегося с преподавателем по видам учебных занятий:		36	-	-	-	36	-	-	-	-
Лекционное занятие (Л)		18	-	-	-	18	-	-	-	-
Семинарское/практическое занятие (СПЗ)		18	-	-	-	18	-	-	-	-
Самостоятельная работа обучающегося, в том числе подготовка к промежуточной аттестации		33	-	-	-	33	-	-	-	-
Вид промежуточной аттестации: Зачет (З), Зачет с оценкой (ЗО), Экзамен (Э), Кандидатский экзамен (КЭ)		<i>Зачет с оценкой</i> 3	-	-	-	30 3	-	-	-	-
Общий объем	в часах	72	-	-	-	72	-	-	-	-
	в зачетных единицах	2	-	-	-	2	-	-	-	-

3. Содержание дисциплины (модуля)

Раздел 1. Введение в статистический анализ биомедицинских данных.

- 1.1. Описательная статистика и разведочный анализ
- 1.2. Проверка статистических гипотез
- 1.3. Кластерный анализ
- 1.4. Обучение с учителем
- 1.5. Корреляция и регрессия
- 1.6. Введение в анализ выживаемости

Раздел 2. Базовые методы системной биологии применительно к анализу омиксных данных

- 2.1 Особенности работы с омиксными данными
- 2.2 Анализ перепредставленности геномных наборов
- 2.3 Основные графовые алгоритмы системной биологии

4. Учебно-тематический план дисциплины (модуля)

Таблица 2

Номер раздела, номер раздела, темы	Наименование разделов, тем	Количество часов					Форма контроля
		Всего	Ауд	Л	СПЗ	СР	
	4 семестр	72	36	18	18	36	Зачет с оценкой
Раздел 1.	Введение в статистический анализ биомедицинских данных.	48	24	12	12	24	Устный опрос
Тема 1.1	Описательная статистика и разведочный анализ	8	4	2	2	4	
Тема 1.2	Проверка статистических гипотез	8	4	2	2	4	
Тема 1.3	Кластерный анализ	8	4	2	2	4	
Тема 1.4	Обучение с учителем	8	4	2	2	4	
Тема 1.5	Корреляция и регрессия	8	4	2	2	4	
Тема 1.6	Введение в анализ выживаемости	8	4	2	2	4	
Раздел 2	Базовые методы системной биологии применительно к анализу омиксных данных.	21	12	6	6	12	
Тема 2.1	Особенности работы с омиксными данными	7	4	2	2	3	
Тема 2.2	Анализ перепредставленности геномных наборов	7	4	2	2	3	
Тема 2.3	Основные графовые алгоритмы системной биологии	7	4	2	2	3	
	Общий объем	72	36	18	18	33	Зачет с оценкой 3

5. Учебно-методическое обеспечение самостоятельной работы обучающихся

Цель самостоятельной работы обучающихся заключается в глубоком, полном усвоении учебного материала и в развитии навыков самообразования. Самостоятельная работа может включать: работу с текстами, литературой, учебно-методическими пособиями, нормативными материалами, в том числе материалами сети интернет, а также проработку конспектов лекций, написание докладов, рефератов, участие в работе семинаров, научных конференциях и пр.

Задания для самостоятельной работы

Таблица 3

Номер раздела	Наименование раздела	Вопросы для самостоятельной работы
1	Введение в статистический анализ биомедицинских данных.	Проведение разведочного анализа данных на языке R и/или Python с применением методов визуализации и кластеризации.
2.	Базовые методы системной биологии применительно к анализу омиксных данных	Реализация метода анализа перепредставленности генных наборов на языке R и/или Python

Контроль самостоятельной работы осуществляется на семинарских (практических) занятиях.

6. Оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся

Примерные варианты оценочных заданий для текущего контроля успеваемости

Таблица 4

Раздел, тема	Наименование разделов, тем	Форма контроля	Оценочное задание
Раздел 1	Введение в статистический анализ биомедицинских данных.	Доклад с защитой презентации по результатам проекта	Выполнить индивидуальный проект по анализу омиксных данных с применением методов статистического анализа и системной биологии.
Раздел 2	Базовые методы системной биологии применительно к анализу омиксных данных.		

Вопросы для подготовки к зачету

1. Описательная статистика.
2. Способы графического изображения выборок: гистограмма, ящики-с-усами, виолин-плот
3. Предобработка и трансформация данных.
4. Проблема пропущенных значений.
5. Проблема выбросов.
6. Проблема эффектов группировки.
7. Анализ главных компонент.
8. Проверка статистических гипотез. Статистические критерии.
9. Таблицы сопряженности. Критерий хи-квадрат.
10. Критерий Стьюдента. Критерий Уилкоксона-Манна-Уитни. Критерий

11. Однофакторный и двухфакторный дисперсионный анализ
12. Проблема множественных сравнений.
13. Линейная регрессия. Методы оценивания адекватности регрессионной модели.
14. Коэффициент корреляции Пирсона и Спирмена.
15. Обучение с учителем. Метрики для оценки двухклассовой классификации.
16. Кросс-валидация.
17. Обучение без учителя (кластерный анализ). Типы расстояний.
18. Иерархическая кластеризация.
19. Разделяющая кластеризация, метод k-средних.
20. Внешние и внутренние индексы оценки качества кластеризации. Оценка числа кластеров.
21. Анализ выживаемости. Функция выживаемости, уровень риска. Оценка Каплан-Мейер.
22. Пропорциональная модель Кокса, допущения и применимость.
23. Особенности омиксных данных.
24. Источники биологических знаний: Gene Ontology, сигнальные каскады, genesets, сети.
25. Анализ перепредставленности.
26. Метод GSEA.
27. Оценка активности сигнальных каскадов с учётом топологии
28. Биологическое уменьшение размерности (сигнальные каскады).
29. Методы поиска мастер-регуляторов.
30. Интеграция различных типов омиксных данных.

Описание критериев и шкал оценивания

В ходе текущего контроля успеваемости (устный или письменный опрос, подготовка и защита реферата, доклад, презентация, тестирование и пр.) при ответах на учебных занятиях, а также промежуточной аттестации в форме кандидатского экзамена обучающиеся оцениваются по четырёхбалльной шкале: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка «отлично» – выставляется аспиранту, если он глубоко усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет связывать теорию с практикой, свободно справляется с задачами и вопросами, не затрудняется с ответами при видоизменении заданий, умеет принять правильное решение и грамотно его обосновывать, владеет разносторонними навыками и приемами выполнения практических задач, комплексной оценкой предложенной ситуации.

Оценка «хорошо» – выставляется аспиранту, если он твердо знает программный материал, грамотно и по существу излагает его, не допуская существенных неточностей при ответе на вопрос, но недостаточно полно раскрывает междисциплинарные связи, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения, комплексной оценкой предложенной ситуации.

Оценка «удовлетворительно» – выставляется аспиранту, если он имеет поверхностные знания программного материала, не усвоил его деталей, допускает неточности, оперирует недостаточно правильными формулировками, нарушает логическую последовательность в изложении программного материала, испытывает затруднения при выполнении практических задач, испытывает затруднения с комплексной оценкой предложенной ситуации, не полностью отвечает на вопросы, в том числе при помощи наводящих вопросов преподавателя.

Оценка «неудовлетворительно» – выставляется аспиранту, который не знает значительной части программного материала, допускает грубые ошибки, неуверенно, с большими затруднениями решает практические задачи или не справляется с ними самостоятельно, не владеет комплексной оценкой ситуации, неверно выбирает тактику действий.

В ходе текущего контроля успеваемости (устный или письменный опрос, подготовка и защита реферата, доклад, презентация, тестирование и пр.) при ответах на учебных занятиях, а также промежуточной аттестации в форме зачета обучающиеся оцениваются по двухбалльной шкале:

Оценка «зачтено» – выставляется аспиранту, если он продемонстрировал знания программного материала, подробно ответил на теоретические вопросы, справился с выполнением заданий и (или) ситуационных задач, предусмотренных рабочей программой дисциплины (модуля).

Оценка «не зачтено» – выставляется аспиранту, если он имеет пробелы в знаниях программного материала, не владеет теоретическим материалом и допускает грубые, принципиальные ошибки в выполнении заданий и (или) ситуационных задач, предусмотренных рабочей программой дисциплины (модуля).

Шкала оценивания (четырёхбалльная или двухбалльная), используемая в рамках текущего контроля успеваемости определяется преподавателем, исходя из целесообразности применения той или иной шкалы.

Если текущий контроль успеваемости и (или) промежуточная аттестация, предусматривает тестовые задания, то перевод результатов тестирования в четырёхбалльную шкалу осуществляется по схеме:

Оценка «Отлично» – 90-100% правильных ответов;

Оценка «Хорошо» – 80-89% правильных ответов;

Оценка «Удовлетворительно» – 71-79% правильных ответов;

Оценка «Неудовлетворительно» – 70% и менее правильных ответов.

Перевод результатов тестирования в двухбалльную шкалу:

Оценка «Зачтено» – 71-100% правильных ответов;

Оценка «Не зачтено» – 70% и менее правильных ответов.

9. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Таблица 5

№ п/п	Автор, наименование, место издания, издательство, год издания	Количество экземпляров
1	Р. Кабаков, R в действии. Анализ и визуализация данных на языке R, Москва, ДМК Пресс, 2016	5
2	С. Мастицкий, В.Шитиков, Статистический анализ и визуализация данных с помощью R, Москва, ДМК Пресс, 2015	5
3	П. Брюс, Практическая статистика для специалистов Data Science. 50+ важнейших понятий с использованием R и Python, Москва, БХВ, 2021	5

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Официальный сайт ИБМХ: адрес ресурса – <https://www.ibmcs.msk.ru>, на котором содержатся сведения об организации и ее подразделениях, локальные нормативные акты, сведения о реализуемых образовательных программах, их учебно-методическом и материально-техническом обеспечении, а также справочная, оперативная и иная информация в рамках электронной образовательной среды. Через официальный сайт обеспечивается доступ всех участников образовательного процесса к различным сервисам и ссылкам.

2. ЭБС ИБМХ – Электронная библиотечная система;
3. ЭБС IPRbooks – Электронно-библиотечная система;
4. ЭБС Айбукс – Электронно-библиотечная система;
5. ЭБС Букап – Электронно-библиотечная система;
6. ЭБС Лань – Электронно-библиотечная система;
7. ЭБС Юрайт – Электронно-библиотечная система.

Перечень профессиональных баз данных и информационных справочных систем

1. <https://www.python.org/> - основной сайт Python
2. <https://www.r-project.org/> - основной сайт R;
3. <https://cran.r-project.org/> - архив пакетов для статистического анализа данных в R;
4. <http://www.bioconductor.org/> Bioconductor – архив пакетов R, предназначенных для анализа молекулярно-биологических данных.
5. <https://www.ncbi.nlm.nih.gov/> NCBI – Национальный центр биотехнологической информации (базы данных и компьютерные ресурсы по биоинформатике);
6. <http://www.ebi.ac.uk/> – базы данных European Bioinformatics Institute (EBI);
7. <http://www.genome.jp/kegg/> KEGG – коллекция баз данных по сигнальным и регуляторным путям Университета Киото;
8. <http://www.uniprot.org/> UniProt – база данных с информацией о белках;

7. Материально-техническое обеспечение дисциплины (модуля)

Таблица 6

№ п/п	Наименование оборудованных учебных аудиторий	Перечень специализированной мебели, технических средств обучения
1	Учебные аудитории для проведения занятий лекционного и семинарского	Компьютерная техника с возможностью подключения к сети "Интернет" и обеспечением доступа к электронной

типов, групповых и индивидуальных консультаций, текущего контроля успеваемости и промежуточной аттестации	образовательной среде ИБМХ.
---	-----------------------------

Программное обеспечение

- Ubuntu 14+ или MS Windows 7,10,11
- Google Chrome, Microsoft Edge, Mozilla Firefox
- 7-Zip
- R
- Python
- Visual Studio Code

10. Методические указания для обучающихся по изучению дисциплины (модуля)

Преподавание дисциплины (модуля) осуществляется в соответствии с Федеральными государственными требованиями.

Основными формами получения и закрепления знаний по данной дисциплине (модулю) являются занятия лекционного и семинарского типа, самостоятельная работа обучающегося, в том числе под руководством преподавателя, прохождение контроля.

Изучение дисциплины (модуля) согласно учебному плану предполагает самостоятельную работу обучающихся. Самостоятельная работа включает в себя изучение литературы, её конспектирование, подготовку к семинарским (практическим) занятиям, текущему контролю успеваемости и промежуточной аттестации.

Текущий контроль успеваемости по дисциплине (модулю) и промежуточная аттестация осуществляются в соответствии с Порядком организации и проведения текущего контроля успеваемости и Порядком проведения промежуточной аттестации обучающихся, устанавливающим формы проведения промежуточной аттестации, ее периодичность и систему оценок.

Наличие электронных образовательных ресурсов позволяет изучать дисциплину (модуль) инвалидам и лицам с ОВЗ.

Особенности изучения дисциплины (модуля) инвалидами и лицами с ОВЗ определены в Положении об организации получения образования для инвалидов и лиц с ограниченными возможностями здоровья.

11. Методические рекомендации преподавателю по организации учебного процесса по дисциплине (модулю)

Преподавание дисциплины (модуля) осуществляется в соответствии с Федеральными государственными требованиями.

При изучении дисциплины (модуля) рекомендуется использовать следующий набор средств и способов обучения:

- рекомендуемую литературу;
- задания для подготовки к семинарам (практическим занятиям) – вопросы для обсуждения и др.;

— задания для текущего контроля успеваемости (задания для самостоятельной работы обучающихся);

— вопросы и задания для подготовки к промежуточной аттестации по итогам изучения дисциплины (модуля).

При проведении занятий лекционного и семинарского типа, в том числе в форме вебинаров и on-line курсов необходимо строго придерживаться учебно-тематического плана дисциплины (модуля), приведенного в разделе 4 данного документа. Необходимо уделить внимание рассмотрению вопросов и заданий, включенных в оценочные задания, при необходимости, решить аналогичные задачи с объяснением алгоритма решения.

Следует обратить внимание обучающихся на то, что для успешной подготовки к текущему контролю успеваемости и промежуточной аттестации нужно изучить литературу, список которой приведен в разделе 7 данной рабочей программы дисциплины (модуля) и иные источники, рекомендованные в подразделах «Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и «Перечень профессиональных баз данных и информационных справочных систем», необходимых для изучения дисциплины (модуля).

Текущий контроль успеваемости и промежуточная аттестация осуществляются в соответствии с Порядком организации и проведения текущего контроля успеваемости и Порядком проведения промежуточной аттестации обучающихся, устанавливающим формы проведения промежуточной аттестации, ее периодичность и систему оценок, с которыми необходимо ознакомить обучающихся на первом занятии.