

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт фундаментальной медицины и биологии
Высшая школа биологии



подписано электронно-цифровой подписью

Программа дисциплины

Статистика и анализ данных

Направление подготовки: 06.04.01 - Биология

Профиль подготовки: Генетика и генетические технологии

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): старший научный сотрудник, к.н. Козлова О.С. (НИЛ Регуляторная геномика, Научный центр Регуляторная геномика), olga-sphinx@yandex.ru

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-1	Готов использовать полученные биологические знания и знания смежных наук, нормативные документы по организации и проведению научно-исследовательских и (или) производственно-технологических работ в профессиональной деятельности в соответствии с профилем программы магистратуры

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- нормативные документы по организации и проведению научно-исследовательских и (или) производственно-технологических работ в области генетики и генетических технологий;
- методы математической статистики и компьютерной биологии для анализа данных в области генетики и генетических технологий;
- требования нормативных документов по организации и проведению научно-исследовательских и (или) производственно-технологических работ в области генетики и генетических технологий.

Должен уметь:

- использовать полученные биологические знания и знания смежных наук для организации и проведения научно-исследовательских и (или) производственно-технологических работ в области генетики и генетических технологий;
- использовать методы математической статистики и компьютерной биологии для анализа данных генетики и генетических технологий;
- определять перечень нормативных документов по организации и проведению научно-исследовательских и (или) производственно-технологических работ в области генетики и генетических технологий.

Должен владеть:

- навыками проведения научно-исследовательских и (или) производственно-технологических работ в области генетики и генетических технологий;
- навыками использования методов математической статистики и компьютерной биологии для анализа данных генетики и генетических технологий;
- навыками организации и проведения научно-исследовательских и (или) производственно-технологических работ в области генетики и генетических технологий согласно требований нормативных документов.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.03 Дисциплины (модули)" основной профессиональной образовательной программы 06.04.01 "Биология (Генетика и генетические технологии)" и относится к части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 1 курсе в 1 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) на 72 часа(ов).

Контактная работа - 28 часа(ов), в том числе лекции - 10 часа(ов), практические занятия - 18 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 44 часа(ов).

Контроль (зачёт / экзамен) - 0 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 1 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)						Самостоятельная работа
			Лекции, всего	Лекции в эл. форме	Практические занятия, всего	Практические в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	
1.	Тема 1. Базовые понятия среды статистической обработки R: переменные, операции, векторы, таблицы	1	2	0	4	0	0	0	2
2.	Тема 2. Сравнение средних	1	2	0	4	0	0	0	10
3.	Тема 3. Корреляция и регрессия	1	2	0	4	0	0	0	10
4.	Тема 4. Анализ номинативных данных	1	2	0	4	0	0	0	10
5.	Тема 5. Многомерный анализ данных	1	2	0	2	0	0	0	12
	Итого		10	0	18	0	0	0	44

4.2 Содержание дисциплины (модуля)

Тема 1. Базовые понятия среды статистической обработки R: переменные, операции, векторы, таблицы

Интерфейс программы Rstudio и основы программирования на языке R. Качественные и количественные переменные в статистике, их примеры в рамках биологического исследования.

Типы переменных в R. Арифметические и логические операции с переменными. Вектор значений переменной как способ представления результатов эксперимента. Представление данных многофакторного эксперимента в виде таблицы типа data frame.

Правила обращения к значениям в таблице. Извлечение заданного набора данных из таблицы (по строкам, по столбцам, по индексу, через условие). Команда subset. Объединение нескольких таблиц в одну (по строкам и по столбцам).

Тема 2. Сравнение средних

Нормальное распределение и распределение Стьюдента, их сходства и различия.

Основные законы нормального распределения. Условия применения теста Стьюдента для сравнения двух средних. Проверка нормальности распределения и гомогенности дисперсий.

Зависимые и независимые выборки. Одновыборочные и двухвыборочные тесты Стьюдента и Уилкоксона. Условия применения дисперсионного анализа (критерия Фишера). Сходства и различия методологий теста Стьюдента и дисперсионного анализа. Критерий Тьюки и множественный тест Уилкоксона как продолжение исследования методом дисперсионного анализа. Поправки на множественное сравнение. Их типы и значение.

Тема 3. Корреляция и регрессия

Линейная корреляция Пирсона и условия её применения. Построение модели линейной регрессии методом наименьших квадратов. Понятия остатков и гомоскедастичности. Функция lm. Непараметрический коэффициент корреляции Спирмена. Тепловые карты корреляций, функция heatmap. Построение матриц корреляций с помощью функции pairs. Логистическая регрессия с количественными и качественными предикторами.

Тема 4. Анализ номинативных данных

Анализ таблиц сопряженности. Тест хи-квадрат для проверки соответствия наблюдаемого распределения признака ожидаемому. Тест хи-квадрат для анализа произвольных таблиц сопряженности и условия его применения. Функция chisq.test. Поправка Йейтса на непрерывность. Точный критерий Фишера для анализа малых выборок.

Тема 5. Многомерный анализ данных

Классификация многомерных методов. Анализ главных компонент. Кластеризация: классификация методов, метод k-means, иерархическая кластеризация. Задача и базовый

алгоритм классификации. ROC-кривые. Деревья принятия решений. Условные операторы if, else и ifelse. Вложенные условные операции. Логические операции "и", "или", "не" в языке R. Алгебра логических операций. Операторы организации циклов while и for, их различия. Команда гер. Применение условных операторов и циклов для преобразования количественных переменных в качественные.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;
- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

R: Анализ и визуализация данных - <http://r-analytics.blogspot.ru/>

Rstudio. Take control of your R code - <https://www.rstudio.com/products/rstudio/>

The R Project for Statistical Computing - <https://www.r-project.org/>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	В ходе лекционных занятий полезно вести конспектирование учебного материала. Необходимо обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации, положительный опыт в ораторском искусстве. Желательно оставить в рабочих конспектах поля, на которых делать пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений. Задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.
практические занятия	В ходе подготовки к семинарам изучить основную литературу, ознакомиться с дополнительной литературой, новыми публикациями в научных изданиях. При этом учесть рекомендации преподавателя и требования учебной программы. Необходимо дорабатывать свой конспект лекции, делая в нем соответствующие записи из литературы, рекомендованной преподавателем и предусмотренной учебной программой. При подготовке к докладу или реферативному сообщению, полезно обращаться за методической помощью к преподавателю и составлять план-конспект своего выступления. Полезно продумать примеры с целью обеспечения тесной связи изучаемой теории с реальной жизнью. Рекомендуется дополнять список использованной литературы современными источниками, не представленными в списке рекомендованной литературы, и в дальнейшем использовать собственные подготовленные учебные материалы при написании курсовых и дипломных работ.
самостоятельная работа	Работу с литературой разумнее начинать с разбора материала, изложенного в лекциях. Для лучшего усвоения материала предпочтительнее после лекции затрачивать 20-30 минут на рассмотрение изложенного материала, отмечая места, вызывающие вопросы или содержащие непонятный текст. Вопросы, которые требуют дополнительного уточнения, можно разобрать, используя учебники или обратившись к преподавателю. С целью углубления знаний по изучаемому вопросу требуется использовать рекомендованную литературу и интернет.
зачет	При подготовке каждый студент должен начать с ознакомления с планом занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованной к данной теме. Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 06.04.01 "Биология" и магистерской программе "Генетика и генетические технологии".

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 06.04.01 - Биология
Профиль подготовки: Генетика и генетические технологии
Квалификация выпускника: магистр
Форма обучения: очное
Язык обучения: русский
Год начала обучения по образовательной программе: 2025

Основная литература:

1. Пирогов, В. Ю. Введение в программирование на языке ассемблера GAS в операционной системе Linux : учебное пособие / В. Ю. Пирогов. - 2-е изд., стер. - Москва : ФЛИНТА, 2024. - 292 с. - ISBN 978-5-9765-5585-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2179288> (дата обращения: 30.06.2025). - Режим доступа: по подписке.
2. Козлов, Н. Н. Математический анализ генетического кода : монография / Н. Н. Козлов. - 4-е изд., электрон. - Москва : Лаборатория знаний, 2023. - 326 с. - ISBN 978-5-93208-617-9. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/297503> (дата обращения: 30.06.2025). - Режим доступа: для авториз. пользователей.
3. Статистическая обработка данных в среде wxMaxima : учебное пособие / Л. А. Коробова, С. Н. Черняева, Ю. А. Сафонова, В. В. Денисенко. - Воронеж : ВГУИТ, 2019. - 64 с. - ISBN 978-5-00032-380-9. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/117818> (дата обращения: 30.06.2025). - Режим доступа: для авториз. пользователей.

Дополнительная литература:

1. Статистика : учебник / В.В. Глинский, Л.К. Серга, В.Г. Ионин [и др.] ; под ред. В.В. Глинского. - 5-е изд., перераб. и доп. - Москва : ИНФРА-М, 2025. - 372 с. - (Среднее профессиональное образование). - ISBN 978-5-16-020348-5. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2169870> (дата обращения: 30.06.2025). - Режим доступа: по подписке.
2. Протасов, Ю. М. Статистика : конспект лекций для студентов заочного отделения / Ю. М. Протасов. - 4-е изд., стер. - Москва : ФЛИНТА, 2024. - 152 с. - ISBN 978-5-9765-0791-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2149795> (дата обращения: 30.06.2025). - Режим доступа: по подписке.
3. Статистика : учебное пособие для вузов / К. Н. Горпинченко, Е. В. Кремянская, А. М. Ляховецкий [и др.]. - 2-е изд., стер. - Санкт-Петербург : Лань, 2024. - 156 с. - ISBN 978-5-507-47762-3. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/415379> (дата обращения: 30.06.2025). - Режим доступа: для авториз. пользователей.

*Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.03 Статистика и анализ данных*

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 06.04.01 - Биология

Профиль подготовки: Генетика и генетические технологии

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "Консультант студента", доступ к которой предоставлен обучающимся. Многопрофильный образовательный ресурс "Консультант студента" является электронной библиотечной системой (ЭБС), предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Полностью соответствует требованиям федеральных государственных образовательных стандартов высшего образования к комплектованию библиотек, в том числе электронных, в части формирования фондов основной и дополнительной литературы.