

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



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ

Проф. Д.А. Таюрский

» _____ 20__ г.

подписано электронно-цифровой подписью

Программа дисциплины

Биостатистика ФТД.Б.1

Направление подготовки: 06.04.01 - Биология

Профиль подготовки: Биохимия, молекулярная биология и биоинформатика

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2019

Автор(ы): Козлова О.С.

Рецензент(ы): Акберова Н.И.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Киямова Р. Г.

Протокол заседания кафедры No ____ от " ____ " _____ 20__ г.

Учебно-методическая комиссия Института фундаментальной медицины и биологии:

Протокол заседания УМК No ____ от " ____ " _____ 20__ г.

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы
2. Место дисциплины в структуре основной профессиональной образовательной программы высшего образования
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
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
 - 6.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы и форм контроля их освоения
 - 6.2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания
 - 6.3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы
 - 6.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций
7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)
 - 7.1. Основная литература
 - 7.2. Дополнительная литература
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 к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

Программу дисциплины разработал(а)(и) научный сотрудник, к.н. Козлова О.С. (НИЛ Экстремальная биология, Научно-клинический центр прецизионной и регенеративной медицины), OISKozlova@kpfu.ru

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

Выпускник, освоивший дисциплину, должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОПК-7	готовностью творчески применять современные компьютерные технологии при сборе, хранении, обработке, анализе и передаче биологической информации для решения профессиональных задач
ПК-3	способностью применять методические основы проектирования, выполнения полевых и лабораторных биологических, экологических исследований, использовать современную аппаратуру и вычислительные комплексы (в соответствии с направленностью (профилем) программы магистратуры)

Выпускник, освоивший дисциплину:

Должен демонстрировать способность и готовность:

- грамотно оформлять, анализировать и использовать данные, полученные в результате проведения биологических экспериментов
- планировать экспериментальную деятельность с учётом возможности дальнейшей статистической обработки её результатов
- использовать современные программные средства визуализации и статистической обработки данных

2. Место дисциплины в структуре основной профессиональной образовательной программы высшего образования

Данная учебная дисциплина включена в раздел "ФТД.Б.1 Факультативные дисциплины" основной профессиональной образовательной программы 06.04.01 "Биология (Биохимия, молекулярная биология и биоинформатика)" и относится к базовой (общепрофессиональной) части.

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 - 36 часа(ов), в том числе лекции - 18 часа(ов), практические занятия - 18 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 36 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	Практические занятия	Лабораторные работы	
1.	Тема 1. Базовые понятия среды статистической обработки R: переменные, операции, векторы	1	2	2	0	4
2.	Тема 2. Работа с таблицами в R	1	2	2	0	4

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)			Самостоятельная работа
			Лекции	Практические занятия	Лабораторные работы	
3.	Тема 3. Описательные статистики	1	2	2	0	4
4.	Тема 4. Визуализация данных в R	1	2	2	0	4
5.	Тема 5. Условные операторы и циклы в R	1	2	2	0	4
6.	Тема 6. Сравнение двух средних: тесты Стьюдента и Уилкоксона	1	2	2	0	4
7.	Тема 7. Дисперсионный анализ	1	2	2	0	4
8.	Тема 8. Корреляция и регрессия	1	2	2	0	4
9.	Тема 9. Анализ номинативных данных: тест хи-квадрат и точный критерий Фишера	1	2	2	0	4
	Итого		18	18	0	36

4.2 Содержание дисциплины

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

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.

Тема 2. Работа с таблицами в R

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 3. Описательные статистики

Выборочное среднее, медиана, дисперсия, стандартное отклонение, размах. Важность правильного выбора описательных статистик. Стандартная ошибка среднего. Агрегированный расчёт описательных статистик для нескольких поднаборов данных (функция aggregate). Задание зависимости переменных в виде формулы. Пропущенные значения и способы их обработки.

Тема 4. Визуализация данных в R

Гистограмма. Функция hist и её параметры. Боксплот, или коробчатый график; его интерпретация. Функция boxplot в R. Диаграмма рассеяния. Общая (generic) функция языка R plot, её параметры. Построение графиков различных типов для встроенных в R наборов экспериментальных данных. Извлечение релевантной биологической информации из графиков.

Тема 5. Условные операторы и циклы в R

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.

Тема 6. Сравнение двух средних: тесты Стьюдента и Уилкоксона

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.

Тема 7. Дисперсионный анализ

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 8. 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 9. Анализ номинативных данных: тест хи-квадрат и точный критерий Фишера

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.

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 образования и науки Российской Федерации от 5 апреля 2017 года №301).

Письмо Министерства образования Российской Федерации №14-55-996ин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений".

Положение от 29 декабря 2018 г. № 0.1.1.67-08/328 "О порядке проведения текущего контроля успеваемости и промежуточной аттестации обучающихся федерального государственного автономного образовательного учреждения высшего образования "Казанский (Приволжский) федеральный университет".

Положение № 0.1.1.67-06/241/15 от 14 декабря 2015 г. "О формировании фонда оценочных средств для проведения текущей, промежуточной и итоговой аттестации обучающихся федерального государственного автономного образовательного учреждения высшего образования "Казанский (Приволжский) федеральный университет".

Положение № 0.1.1.56-06/54/11 от 26 октября 2011 г. "Об электронных образовательных ресурсах федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет".

Регламент № 0.1.1.67-06/66/16 от 30 марта 2016 г. "Разработки, регистрации, подготовки к использованию в учебном процессе и удаления электронных образовательных ресурсов в системе электронного обучения федерального государственного автономного образовательного учреждения высшего образования "Казанский (Приволжский) федеральный университет".

Регламент № 0.1.1.67-06/11/16 от 25 января 2016 г. "О балльно-рейтинговой системе оценки знаний обучающихся в федеральном государственном автономном образовательном учреждении высшего образования "Казанский (Приволжский) федеральный университет".

Регламент № 0.1.1.67-06/91/13 от 21 июня 2013 г. "О порядке разработки и выпуска учебных изданий в федеральном государственном автономном образовательном учреждении высшего профессионального образования "Казанский (Приволжский) федеральный университет".

6. Фонд оценочных средств по дисциплине (модулю)

6.1 Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы и форм контроля их освоения

Этап	Форма контроля	Оцениваемые компетенции	Темы (разделы) дисциплины
Семестр 1			
	Текущий контроль		
1	Компьютерная программа	ПК-2, ПК-1	5. Условные операторы и циклы в R 6. Сравнение двух средних: тесты Стьюдента и Уилкоксона 7. Дисперсионный анализ 8. Корреляция и регрессия 9. Анализ номинативных данных: тест хи-квадрат и точный критерий Фишера
2	Устный опрос	ПК-2	1. Базовые понятия среды статистической обработки R: переменные, операции, векторы 2. Работа с таблицами в R 3. Описательные статистики 4. Визуализация данных в R
	Зачет	ОПК-7, ПК-3	

6.2 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Форма контроля	Критерии оценивания				Этап
	Отлично	Хорошо	Удовл.	Неуд.	
Семестр 1					
Текущий контроль					
Компьютерная программа	Высокий уровень умений и навыков программирования, в том числе моделирования, алгоритмизации, использования языка программирования. Поставленная задача полностью решена.	Хороший уровень умений и навыков программирования, в том числе моделирования, алгоритмизации, использования языка программирования. Поставленная задача в основном решена.	Удовлетворительный уровень умений и навыков программирования, в том числе моделирования, алгоритмизации, использования языка программирования. Поставленная задача решена частично.	Недостаточный уровень умений и навыков программирования, в том числе моделирования, алгоритмизации, использования языка программирования. Поставленная задача не решена.	1
Устный опрос	В ответе качественно раскрыто содержание темы. Ответ хорошо структурирован. Прекрасно освоен понятийный аппарат. Продемонстрирован высокий уровень понимания материала. Превосходное умение формулировать свои мысли, обсуждать дискуссионные положения.	Основные вопросы темы раскрыты. Структура ответа в целом адекватна теме. Хорошо освоен понятийный аппарат. Продемонстрирован хороший уровень понимания материала. Хорошее умение формулировать свои мысли, обсуждать дискуссионные положения.	Тема частично раскрыта. Ответ слабо структурирован. Понятийный аппарат освоен частично. Понимание отдельных положений из материала по теме. Удовлетворительное умение формулировать свои мысли, обсуждать дискуссионные положения.	Тема не раскрыта. Понятийный аппарат освоен неудовлетворительно. Понимание материала фрагментарное или отсутствует. Неумение формулировать свои мысли, обсуждать дискуссионные положения.	2
	Зачтено		Не зачтено		
Зачет	Обучающийся обнаружил знание основного учебно-программного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по специальности, справился с выполнением заданий, предусмотренных программой дисциплины.		Обучающийся обнаружил значительные пробелы в знаниях основного учебно-программного материала, допустил принципиальные ошибки в выполнении предусмотренных программой заданий и не способен продолжить обучение или приступить по окончании университета к профессиональной деятельности без дополнительных занятий по соответствующей дисциплине.		

6.3 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Семестр 1

Текущий контроль

1. Компьютерная программа

Темы 5, 6, 7, 8, 9

По каждой теме каждому студенту предоставляется индивидуальная задача, для решения которой необходимо написать небольшую программу в среде статистических вычислений R. Кроме того, каждый студент в устной форме защищает свою программу, мотивируя выбор метода статистической обработки и интерпретируя полученные результаты.

Тема 5.

1. Дан код "if(a>0) {a<-a-5} else {a<-a+5}". Допустим, a равно 5. Что произойдет после первого запуска приведённой команды и после второго её запуска?

2. Дан код "if(a>0) {a<-a-5} else if(a<0) {a<-a+5} else {a<-a+0}". Допустим, a равно 5. Что произойдет после первого запуска приведённой команды и после второго её запуска?

3. Дана таблица, хранящая оценки, выставяемые студентами своим преподавателям. Для преподавателей известны возраст и другая информация. Создать в таблице, хранящей данные оценок для преподавателей, ещё один столбец - `age_group`, хранящий "young", если возраст преподавателя меньше 40, "mature", если возраст преподавателя не меньше 40, но меньше 60, и "senior" во всех остальных случаях. Сколько человек принадлежит к каждой категории?
4. Дан код: `"a<-1; for (i in 1:100) {a<-a+1}"`. Чему будет равно `a` по окончании работы кода?
5. Дан код: `"a<-1:10; for (i in 1:10) {a[i]<-a[i]+i}"`. Чему будет равно максимальное значение в данном векторе?
6. Создать в таблице с оценками преподавателей новую переменную - `quality_for_age`, которая будет хранить "good", если оценка данного преподавателя выше средней оценки преподавателей с таким же возрастом, и "bad" в противных случаях.
7. Дан код: `"a<-2; while(a<5) {a<-a+1}"`. Сколько раз была выполнена команда внутри фигурных скобок?
8. Продвигаясь последовательно по строкам таблицы с оценками преподавателей, суммировать оценки, пока сумма меньше 100. На какой строке сумма превысит 100 баллов?
9. Используя встроенный в R набор данных `chickwts`, содержащий информацию о весе цыплят в граммах, выкормленных на разных кормах, составить таблицу описательных статистик.
10. Используя встроенный в R набор данных `chickwts`, сравнить внутригрупповые средние веса цыплят и средний вес без учёта деления на группы (в граммах). Определить, у цыплят с каким типом откорма средний внутригрупповой вес превышает общий средний, а с каким - является меньше общего среднего. Добавить в таблицу описательных статистик столбец "Greater than mean", который будет содержать "Yes", если средний вес цыплят в данной группе больше общего среднего, и "No", в противном случае.

Тема 6.

1. Исследовать влияние витамина C и апельсинового сока в дозировке 0.5 мг на рост зубов морских свинок (встроенный в R набор данных `ToothGrowth`).
2. Используя встроенный в R набор данных `chickwts`, построить коробчатый график для веса цыплят и определить, какому типу корма соответствует цыплёнок с минимальным и максимальным весом. Охарактеризовать распределение весов цыплят для типа корма, которому принадлежит цыплёнок с максимальным весом.
3. Используя встроенный в R набор данных `CO2`, содержащий информацию о темпах поглощения двуокиси углерода растением *Echinochloa crus-galli* (ежовник обыкновенный), произрастающем в Квебеке и Миссисипи, в зависимости от концентрации `CO2` в окружающем воздухе и того факта, было ли растение охлаждено накануне проведения эксперимента или нет, составить таблицу описательных статистик.
4. Построить коробчатый график для зависимости темпов поглощения двуокиси углерода от фактора происхождения растения (`Type`) и того факта, было ли оно предварительно охлаждено (`Treatment`). Охарактеризовать зависимость темпов поглощения `CO2` от этих двух факторов и определить, какой из них, предположительно, определяет большую изменчивость признака.
5. Построить коробчатый график для зависимости темпов поглощения двуокиси углерода от его концентрации (`conc`) и охарактеризовать данную зависимость.
6. Найти растения с минимальным и максимальным средним темпом поглощения двуокиси углерода.
7. Заменить в своей копии таблицы `CO2` значения в столбце `Type` на `tolerant/non-tolerant` (`tolerant` - для растений, менее чувствительных к холоду, чем растения `non-tolerant`).
8. Используя команду `source` ("<http://www.openintro.org/stat/data/present.R>"), сохранить фрейм данных под названием `present` с сайта `OpenIntro`. Этот массив данных содержит количества рождений мальчиков и девочек с 1940 по 2002 год в США. Рассчитайте абсолютные различия между количеством мальчиков и девочек, родившихся в каждом году, и определите, в каком году была самая большая абсолютная разница в количествах новорожденных девочек и мальчиков?
9. На одном графике отразите динамику рождений мальчиков и девочек. На основе графика определите, справедливо ли утверждение, что каждый год рождалось больше девочек, чем мальчиков.
10. Постройте график доли мальчиков с течением времени. На основе графика определите, справедливо ли утверждение, что доля мальчиков, родившихся в США, уменьшилась с течением времени.

Тема 7.

1. Исследовать влияние вида ириса на длину лепестка (встроенный в R набор данных `iris`).
2. Исследовать влияние вида ириса на ширину лепестка (встроенный в R набор данных `iris`).
3. Исследовать влияние вида ириса на длину чашелистика (встроенный в R набор данных `iris`).
4. Исследовать влияние вида ириса на ширину чашелистика (встроенный в R набор данных `iris`).
5. Используя команду `source` (<http://www.openintro.org/stat/data/cdc.R>), загрузите в рабочее пространство базу данных `cdc`, в которой представлены переменные: `genhlth`, `exerany`, `hlthplan`, `smoke100`, `height`, `weight`, `wtdesire`, `age`, и `gender`. Создайте сводную таблицу для всех переменных фрейма данных `cdc`. Какие переменные являются количественными? Распределение каких признаков является нормальным?
6. Посчитайте относительное распределение частот для `genhlth`. Какую долю составляют респонденты, находящиеся в отличном здоровье?
7. Каково количество курящих и некурящих респондентов каждого пола? Представьте результат на мозаичной диаграмме. Что можно выявить о привычке к курению среди мужчин и женщин по мозаичной диаграмме?
8. Создайте новый объект с именем `under23_and_smoke`, который содержит все наблюдения для респондентов в возрасте до 23, которые выкурили хотя бы 100 сигарет. Запишите команду, что вы использовали для создания нового объекта. Сколько наблюдений в созданном объекте `under23_and_smoke`?

9. Построить коробчатый график для зависимости веса (weight) от состояния здоровья (genhlth) респондентов. Какие предположения можно сделать на основе этого графика?

10. Постройте график зависимости веса (weight) от желаемого веса (wt desire) и охарактеризуйте связь между этими переменными.

Тема 8.

1. Построить диаграмму рассеяния для зависимости расхода топлива автомобилей из встроенного набора данных mtcars от количества лошадиных сил.

2. Рассчитайте коэффициенты корреляции Пирсона и Спирмена между количеством лошадиных сил и расходом топлива (набор данных mtcars).

3. Визуализируйте все возможные парные диаграммы рассеяния между количественными параметрами набора данных mtcars (на одном графике). Между какими параметрами есть очевидная зависимость?

4. Составьте уравнение линейной регрессии, описывающее зависимость расхода топлива от количества лошадиных сил. Добавьте на диаграмму рассеяния красную линию, соответствующую уравнению регрессии.

5. Используя составленное уравнение регрессии, предскажите расход топлива для новых наблюдений в таблице (100, 150, 129, 300 лошадиных сил).

6. Составьте уравнение линейной регрессии для зависимости расхода топлива от факторной переменной - числа цилиндров.

7. Построить матрицу диаграмм рассеяния между всеми парами измерений из набора встроенных в R данных iris.

8. Используя команду source (<http://www.openintro.org/stat/data/cdc.R>), загрузите в рабочее пространство базу данных cdc, в которой представлены переменные: genhlth, exerany, hlthplan, smoke100, height, weight, wt desire, age, и gender. Добавьте новую переменную индекс массы тела (BMI) во фрейме данных cdc. Как зависит BMI от состояния здоровья (genhlth) респондентов? Отобразите результат на графике.

9. Загрузите базу данных nc в рабочее пространство RStudio с http://d396qusza40orc.cloudfront.net/statistics/lab_resources/nc.RData. Связан ли вес ребенка (weight) с весом, набранным матерью во время беременности (gained)? Постройте график зависимости и охарактеризуйте связь между этими переменными.

10. Связан ли вес ребенка (weight) с возрастом матерей? Постройте график зависимости и охарактеризуйте связь между этими переменными.

Тема 9.

1. Дана таблица сопряженности курения и наличия артериальной гипертонии. Исследовать взаимосвязь между этими двумя факторами.

2. Даны наблюдаемые частоты расщепления по фенотипу в F2 дигибридного скрещивания. Проверить, соответствует ли оно ожидаемому.

3. Дана таблица сопряженности между типом анестезии (галотан/морфин) и операционной летальностью. Исследовать достоверность этой взаимосвязи.

4. Дана таблица сопряженности между неким фактором 1 с тремя градациями и фактором 2 с двумя. Проверить нулевую гипотезу об отсутствии взаимосвязи этих факторов с помощью критерия хи-квадрат.

5. Рассчитать точную вероятность случайно получить заданную или ещё более несбалансированную таблицу сопряженности между неким фактором 1 и фактором 2 (оба с двумя градациями).

6. В лаборатории работает 5 юношей и 8 девушек. Можно ли сказать, что такое распределение статистически значимо отличается от равномерного?

7. Изучалась эффективность высокочастотной стимуляции нерва в качестве обезболивающего средства при удалении зуба. Все больные подключались к прибору, но в одних случаях он работал, в других был выключен. Ни стоматолог, ни больной не знали, включён ли прибор. Дана таблица сопряженности. Позволяют ли эти данные считать высокочастотную стимуляцию нерва действенным анальгезирующим средством?

8. Составить уравнение логистической регрессии для данных о студентах (дана таблица сопряженности сдачи/несдачи экзамена и наличия и отсутствия подготовки).

9. Мендель выращивал горох трёх цветов. В одном из опытов эмпирическое распределение частот некоторого цвета гороха приняло следующий вид: 18, 55, 27. Предполагаемое теоретическое распределение имеет следующий вид: 1:2:1. Рассчитайте расстояние хи-квадрат для этого примера.

10. Одна из причин инсульта - окклюзия сонной артерии. Чтобы выяснить, какое лечение - медикаментозное или хирургическое - даёт в этом случае лучшие результаты, исследователи сравнили долгосрочный прогноз у пациентов, на которых применялись два этих метода. Дана таблица сопряженности. Можно ли говорить о превосходстве одного из видов лечения?

2. Устный опрос

Темы 1, 2, 3, 4

Тема 1.

1. Как создать переменную в R и присвоить ей значение?

2. Какие существуют правила при выборе имени переменной?

3. Какие логические операции вы знаете?

4. Что такое вектор в R и как его создать?

5. Правила сложения векторов.

6. Типы переменных в R

7. Правила и возможные способы преобразования типов
8. Программа Rstudio, её окна и основные вкладки
9. Что такое скрипт на языке R? Как его создать и запустить?
10. Как установить рабочую директорию в программе Rstudio?

Тема 2.

1. Как определить, сколько столбцов и строк имеет таблица в R?
2. Какие функции для получения информации о наборе данных (data frame) вы знаете?
3. Что такое факторная переменная в R, в чём её отличие от простой символьной переменной? 4. Какие существуют способы извлечения поднабора данных из большой таблицы?
5. Как объединить несколько таблиц в одну?
6. Правила работы с именами строк и столбцов таблицы
7. Как загрузить и выгрузить таблицу?
8. Какие разделители столбцов вы знаете? Как указать, какой именно разделитель используется?
9. Как отсортировать таблицу в режиме просмотра в программе Rstudio?
10. Добавление и удаление столбцов в data frame

Тема 3.

1. Что такое описательные статистики и зачем они нужны?
2. Какие описательные статистики вы знаете?
3. Классификация описательных статистик
4. Функции, соответствующие описательным статистикам, в R
5. Генеральная и выборочная дисперсия, их различия
6. Связь между стандартным отклонением и стандартной ошибкой среднего
7. Как быстро рассчитать описательные статистики сразу для нескольких поднаборов данных? 8. Чем отличается тип данных "список" (list) от вектора?
9. Как задать зависимость между переменными в R?
10. Понятие нормального распределения

Тема 4.

1. Что такое гистограмма и для чего она используется?
2. Как построить гистограмму в R? Какие параметры имеет функция гистограммы?
3. Какие задачи решаются с помощью построения коробчатых графиков?
4. Что такое выбросы и как с ними обращаться?
5. Что отражают "усы" графика типа boxplot?
6. Как построить boxplot для одной (нескольких) переменных в R?
7. Что такое gencis функции в R? Приведите примеры таких функций
8. Диаграмма рассеяния и параметры её построения
9. Функция для построения матриц диаграмм рассеяния в R
10. Основные функции пакета ggplot2

Зачет

Вопросы к зачету:

1. Основные типы переменных для представления результатов биологических экспериментов
2. Описательные статистики
3. Нормальное распределение, его свойства
4. Формальные статистические тесты на нормальность распределения и гомогенность дисперсий
5. Доверительный интервал для среднего по нормальному и t-распределению
6. Статистические гипотезы, ошибки и понятие статистического вывода
7. Одновыборочные и двухвыборочные тесты Стьюдента и Уилкоксона
8. Сравнение средних для зависимых и независимых выборок
9. Понятие дисперсионного анализа
10. Попарные сравнения средних. Поправка Бонферрони и критерий Тьюки
11. Непараметрические тесты для сравнения средних между множественными выборками
12. Коэффициент корреляции Пирсона. Коэффициент детерминации
13. Понятие регрессионного анализа
14. Понятие логистической регрессии
15. Критерий хи-квадрат
16. Точный критерий Фишера
17. Коробчатый график. Его построение и интерпретация
18. Гистограмма как способ визуализации экспериментальных данных
19. Диаграмма рассеяния и способы её использования
20. Базовый протокол разведочного анализа данных

6.4 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

В КФУ действует балльно-рейтинговая система оценки знаний обучающихся. Суммарно по дисциплине (модулю) можно получить максимум 100 баллов за семестр, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов.

Для зачёта:

56 баллов и более - "зачтено".

55 баллов и менее - "не зачтено".

Для экзамена:

86 баллов и более - "отлично".

71-85 баллов - "хорошо".

56-70 баллов - "удовлетворительно".

55 баллов и менее - "неудовлетворительно".

Форма контроля	Процедура оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций	Этап	Количество баллов
Семестр 1			
Текущий контроль			
Компьютерная программа	Обучающиеся самостоятельно составляют программу на определённом языке программирования в соответствии с заданием. Программа сдаётся преподавателю в электронном виде. Оценивается реализация алгоритмов на языке программирования, достижение заданного результата.	1	30
Устный опрос	Устный опрос проводится на практических занятиях. Обучающиеся выступают с докладами, сообщениями, дополнениями, участвуют в дискуссии, отвечают на вопросы преподавателя. Оценивается уровень домашней подготовки по теме, способность системно и логично излагать материал, анализировать, формулировать собственную позицию, отвечать на дополнительные вопросы.	2	20
Зачет	Зачёт нацелен на комплексную проверку освоения дисциплины. Обучающийся получает вопрос (вопросы) либо задание (задания) и время на подготовку. Зачёт проводится в устной, письменной или компьютерной форме. Оценивается владение материалом, его системное освоение, способность применять нужные знания, навыки и умения при анализе проблемных ситуаций и решении практических заданий.		50

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

7.1 Основная литература:

1. Бослаф С., Статистика для всех / Сара Бослаф - М. : ДМК Пресс, 2015. - 586 с. - ISBN 978-5-94074-969-1 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785940749691.html> (дата обращения: 19.06.2019). - Режим доступа : по подписке.

2. Шипунов А.Б., Наглядная статистика. Используем R! / А.Б. Шипунов, Е.М. Балдин, П.А. Волкова, А.И. Коробейников, С.А. Назарова, С.В. Петров, В.Г. Суфиянов. - М. : ДМК Пресс, 2012. - 298 с. - ISBN 978-5-94074-828-1 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785940748281.html> (дата обращения: 19.06.2019). - Режим доступа : по подписке.

3. Лагутин, М. Б. Наглядная математическая статистика : учебное пособие / М. Б. Лагутин. - 7-е изд. - Москва : Лаборатория знаний, 2019. - 475 с. - ISBN 978-5-00101-642-7. - Текст : электронный // Лань : электронно-библиотечная система. - URL : <https://e.lanbook.com/book/116104> (дата обращения: 19.06.2019). - Режим доступа: для авториз. пользователей.

7.2. Дополнительная литература:

1. Леонов С.А., Статистические методы анализа в здравоохранении. Краткий курс лекций / Леонов С.А., Вайсман Д.Ш., Моравская С.В, Мирсков Ю.А. - М. : Менеджер здравоохранения, 2011. - 172 с. - ISBN 978-5-903834-11-2 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785903834112.html> (дата обращения: 19.06.2019). - Режим доступа : по подписке.

2. Кучеренко В.З., Применение методов статистического анализа для изучения общественного здоровья и здравоохранения : учебное пособие / Под ред. В.З. Кучеренко. - 4 изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. - ISBN 978-5-9704-1915-1 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785970419151.html> (дата обращения: 19.06.2019). - Режим доступа : по подписке.

3. Неделько С.В., Типовые задачи математической статистики : учеб. пособие / Неделько С.В. - Новосибирск : Изд-во НГТУ, 2014. - 52 с. - ISBN 978-5-7782-2481-0 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785778224810.html> (дата обращения: 19.06.2019). - Режим доступа : по подписке.

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. Примите к сведению следующие рекомендации: Обращайте 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	С самого начала изучения курса следует регулярно работать с компьютерной средой автоматизированного анализа данных (R, Rstudio). Изучение курса предполагает использование учебных пособий (в том числе электронных), рабочих тетрадей по курсу, базовых рекомендованных учебников (в том числе электронных), а также обзоров, доступных в интернет ресурсах, рекомендованных преподавателем.
самостоятельная работа	Самостоятельная работа студентов (СРС) включает следующие виды работ: - проработка материала, изучаемого на практикуме (типовые алгоритмы решения задач) - подготовка к контрольным работам (решению индивидуальных задач посредством написания программного кода) Рекомендуется установка приложений R и Rstudio на домашний ПК студента.
компьютерная программа	При написании кода на языке R студент должен учитывать рекомендации преподавателя и отталкиваться от решений типовых задач, разбираемых на практических занятиях. Рекомендуется использовать содержательные названия переменных, а также соблюдать чистоту кода. В программе не должно быть ненужных переменных, а все комментарии перед сдачей программы преподавателю должны быть удалены.
устный опрос	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. Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы семинара, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы.
зачет	Зачёт нацелен на комплексную проверку освоения дисциплины. Обучающийся получает вопрос (вопросы) либо задание (задания) и время на подготовку. Зачёт проводится в устной, письменной или компьютерной форме. Оценивается владение материалом, его системное освоение, способность применять нужные знания, навыки и умения при анализе проблемных ситуаций и решении практических заданий.

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

Освоение дисциплины "Биостатистика" предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows Professional 7 Russian

Пакет офисного программного обеспечения Microsoft Office 2010 Professional Plus Russian

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI

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.

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

Освоение дисциплины "Биостатистика" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

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 "Биохимия, молекулярная биология и биоинформатика".