

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. Таюрский Д.А.

_____ 20__ г.

Программа дисциплины

Практикум по биostatистике и анализу данных Б1.В.ОД.4.2

Направление подготовки: 06.03.01 - Биология

Профиль подготовки: не предусмотрено

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Козлова О.С., Тризна Е.Ю., Фахруллин Р.Ф.

Рецензент(ы):

Чернов В.М.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Киямова Р. Г.

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института фундаментальной медицины и биологии:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

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; ассистент, б.с. Тризна Е.Ю. кафедра генетики Центр биологии и педагогического образования, trizna91@mail.ru; главный научный сотрудник, д.н. Фахруллин Р.Ф. НИЛ OpenLab Бионанотехнологии Центр научной деятельности и аспирантуры, kazanbio@gmail.com

1. Цели освоения дисциплины

Основной целью курса является ознакомление студентов с практическим применением теории вероятности и математической статистики. Задача дисциплины - правильно планировать и выполнять научные исследования, применяя методы математической статистики, проводя статистическую обработку данных и на основе полученных расчетов делать выводы.

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

Данная учебная дисциплина включена в раздел "Б1.В.ОД.4 Дисциплины (модули)" основной образовательной программы 06.03.01 Биология и относится к обязательным дисциплинам. Осваивается на 2 курсе, 3 семестр.

Данная дисциплина относится к циклу 'Математические и естественнонаучные дисциплины' всех профилей подготовки по направлению подготовки 020400 - Биология.

Математические методы в биологии - комплексное научное направление, имеющее междисциплинарный характер. Студент должен иметь базовые понятия в области математики, логики, биологии, а также навыки работы на персональном компьютере.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОПК-1 (профессиональные компетенции)	способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности
ПК-4 (профессиональные компетенции)	способность применять современные методы обработки, анализа и синтеза полевой, производственной и лабораторной биологической информации, правила составления научно-технических проектов и отчетов

В результате освоения дисциплины студент:

1. должен знать:

основные законы теории вероятностей, комбинаторики и математической статистики, возможности современных программных средств.

2. должен уметь:

работать в качестве пользователя в статистических пакетах, самостоятельно формулировать задачи математической статистики и правильно применять различные статистические критерии.

3. должен владеть:

методами математической статистики для описания полученных данных, проверки достоверности полученных результатов.

4. 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

4. Структура и содержание дисциплины/ модуля

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

Форма промежуточного контроля дисциплины: зачет в 3 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практи- ческие занятия	Лабора- торные работы	
1.	Тема 1. Базовые понятия среды статистической обработки R: переменные, операции, векторы	3		0	4	0	Устный опрос
2.	Тема 2. Работа с таблицами в R	3		0	4	0	Устный опрос
3.	Тема 3. Описательные статистики	3		0	4	0	Устный опрос
4.	Тема 4. Визуализация данных в R	3		0	4	0	Устный опрос
5.	Тема 5. Условные операции и циклы в R	3		0	4	0	Компьютерная программа
6.	Тема 6. Сравнение двух средних: тесты Стьюдента и Уилкоксона	3		0	4	0	Компьютерная программа
7.	Тема 7. Дисперсионный анализ	3		0	4	0	Компьютерная программа

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практи- ческие занятия	Лабора- торные работы	
8.	Тема 8. Корреляция и регрессия	3		0	4	0	Компьютерная программа
9.	Тема 9. Анализ номинативных данных: тест хи-квадрат и точный критерий Фишера	3		0	4	0	Компьютерная программа
.	Тема . Итоговая форма контроля	3		0	0	0	Зачет
	Итого			0	36	0	

4.2 Содержание дисциплины

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

практическое занятие (4 часа(ов)):

Интерфейс программы Rstudio. Качественные и количественные переменные в статистике. Арифметические и логические операции с переменными. Вектор значений переменной как способ представления результатов эксперимента.

Тема 2. Работа с таблицами в R

практическое занятие (4 часа(ов)):

Представление данных в виде таблицы типа data frame. Извлечение заданного набора данных из таблицы. Объединение нескольких таблиц в одну.

Тема 3. Описательные статистики

практическое занятие (4 часа(ов)):

Выборочное среднее, медиана, стандартное отклонение, размах. Агрегированный расчёт описательных статистик для нескольких поднаборов данных. Задание зависимости переменных в виде формулы. Пропущенные значения и способы их обработки.

Тема 4. Визуализация данных в R

практическое занятие (4 часа(ов)):

Гистограмма. Боксплот, или коробчатый график. Диаграмма рассеяния.

Тема 5. Условные операции и циклы в R

практическое занятие (4 часа(ов)):

Условный оператор if. Операторы организации циклов while и for. Применение условных операторов и циклов для преобразования количественных переменных в качественные.

Тема 6. Сравнение двух средних: тесты Стьюдента и Уилкоксона

практическое занятие (4 часа(ов)):

Нормальное распределение и распределение Стьюдента. 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

практическое занятие (4 часа(ов)):

Условия применения дисперсионного анализа. Критерий Тьюки и множественный тест Уилкоксона. Поправки на множественное сравнение.

Тема 8. Корреляция и регрессия

практическое занятие (4 часа(ов)):

Линейная корреляция Пирсона и условия её применения. Построение модели линейной регрессии. Непараметрический коэффициент корреляции Спирмена. Логистическая регрессия с количественными и качественными предикторами.

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

практическое занятие (4 часа(ов)):

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

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Базовые понятия среды статистической обработки R: переменные, операции, векторы	3		подготовка к устному опросу	4	устный опрос
2.	Тема 2. Работа с таблицами в R	3		подготовка к устному опросу	4	устный опрос
3.	Тема 3. Описательные статистики	3		подготовка к устному опросу	4	устный опрос
4.	Тема 4. Визуализация данных в R	3		подготовка к устному опросу	4	устный опрос
5.	Тема 5. Условные операции и циклы в R	3		создание компьютерной программы	4	компьютерная программа
6.	Тема 6. Сравнение двух средних: тесты Стьюдента и Уилкоксона	3		создание компьютерной программы	4	компьютерная программа
7.	Тема 7. Дисперсионный анализ	3		создание компьютерной программы	4	компьютерная программа
8.	Тема 8. Корреляция и регрессия	3		создание компьютерной программы	4	компьютерная программа

N	Раздел дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
9.	Тема 9. Анализ номинативных данных: тест хи-квадрат и точный критерий Фишера	3		создание компьютерной программы	4	компьютерная программа
	Итого				36	

5. Образовательные технологии, включая интерактивные формы обучения

На практических занятиях студенты решают типовые задачи по статистическому анализу результатов биологических экспериментов на компьютере, с использованием языка программирования R и программы Rstudio.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

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

устный опрос, примерные вопросы:

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. Работа с таблицами в R

устный опрос, примерные вопросы:

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. Описательные статистики

устный опрос, примерные вопросы:

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. Визуализация данных в R

устный опрос, примерные вопросы:

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 generic 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

Тема 5. Условные операции и циклы в R

компьютерная программа, примерные вопросы:

1. Дан код "if(a>0) {a<-a-5} else {a<-a+5}". Допустим, a равно 5. Что произойдёт 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 произойдёт после первого запуска приведённой команды и после второго её запуска? 3. Дана таблица, хранящая оценки, выставляемые 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. Сравнение двух средних: тесты Стьюдента и Уилкоксона

компьютерная программа, примерные вопросы:

1. Исследовать влияние витамина С 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. Дисперсионный анализ

компьютерная программа, примерные вопросы:

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, wt desire, 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. Корреляция и регрессия

компьютерная программа, примерные вопросы:

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`, `wtdesire`, `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. Анализ номинативных данных: тест хи-квадрат и точный критерий Фишера

компьютерная программа, примерные вопросы:

1. Дана таблица сопряженности курения и наличия артериальной гипертензии. 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?

Итоговая форма контроля

зачет (в 3 семестре)

Примерные вопросы к итоговой форме контроля

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

7.1. Основная литература:

1. Статистика для всех [Электронный ресурс] / Сара Бослаф - М. : ДМК Пресс, 2015. - <http://www.studentlibrary.ru/book/ISBN9785940749691.html>
2. Наглядная статистика. Используем R! [Электронный ресурс] / 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. - <http://www.studentlibrary.ru/book/ISBN9785940748281.html>

7.2. Дополнительная литература:

1. Статистические методы анализа в здравоохранении. Краткий курс лекций [Электронный ресурс] / Леонов С.А., Вайсман Д.Ш., Моравская С.В, Мирсков Ю.А. - М. : Менеджер здравоохранения, 2011. - <http://www.studentlibrary.ru/book/ISBN9785903834112.html>
2. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс] : учебное пособие / Под ред. В.З. Кучеренко. - 4 изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - <http://www.studentlibrary.ru/book/ISBN9785970419151.html>
3. Типовые задачи математической статистики [Электронный ресурс]: учеб. пособие / Неделько С.В. - Новосибирск : Изд-во НГТУ, 2014. - <http://www.studentlibrary.ru/book/ISBN9785778224810.html>

7.3. Интернет-ресурсы:

- Learn R Programming - <https://www.programiz.com/r-programming>
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/>
Try R - <http://tryr.codeschool.com/>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Практикум по биостатистике и анализу данных" 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.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен студентам. 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 студентам. Электронная библиотечная система "Консультант студента" предоставляет полнотекстовый доступ к современной учебной литературе по основным дисциплинам, изучаемым в медицинских вузах (представлены издания как чисто медицинского профиля, так и по естественным, точным и общественным наукам). ЭБС предоставляет вузу наиболее полные комплекты необходимой литературы в соответствии с требованиями государственных образовательных стандартов с соблюдением авторских и смежных прав.

Свободно распространяемая система статистической обработки данных R и её пользовательский интерфейс Rstudi

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 06.03.01 "Биология" и профилю подготовки не предусмотрено.

Автор(ы):

Фахруллин Р.Ф. _____

Козлова О.С. _____

Тризна Е.Ю. _____

"__" _____ 201__ г.

Рецензент(ы):

Чернов В.М. _____

"__" _____ 201__ г.