**Программа учебной дисциплины «Программирование на языке R»**

Утверждена

Академическим советом ООП

Протокол № от «\_\_»\_\_\_\_\_20\_\_ г.

|  |  |
| --- | --- |
| Автор  | Мазин Павел Владимирович |
| Число кредитов  | 4 |
| Контактная работа (час.)  |  |
| Самостоятельная работа (час.)  |  |
| Курс  | 1 |
| Формат изучения дисциплины | без использования онлайн курса |

1. **ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ**

Язык R — современный язык программирования предназначенный для статистического анализа большого объёма данных. На данный момент R является одним из самых используемых языков программирования в биоинформатике.

Цель курса - освоить базовый синтаксис языка R, познакомиться с структурами данных, управляющими конструкциями и графическими возможностями. Курс включает в себя знакомство с функциями R предназначенными для выполнения основных статистических тестов, освоение базовой графики и знакомство с некоторыми биоинформатическими пакетами из Bioconductor.

В результате освоения дисциплины студент должен:

**знать:**

- типы данных в языке R

- базовый синтаксис R

- основные функции предназначенные для анализа данных

**уметь:**

- загрузить данные из файла и сохранить результаты анализа в файл

- провести статистический анализ данных с использованием таких методов как т-тест, тест Вилкоксона, линейные модели и др.

- визуализировать результаты анализа при помощи базовой графики R

**владеть:**

- навыками программирования на языке R

- методами статистического анализа биологических данных

- навыками представления результатов анализа

Изучение дисциплины «Программирование на языке R» базируется на следующих дисциплинах:

- базовое владение персональным компьютером

- базовое знакомство с математическим анализом и математической статистикой

Основные положения дисциплины должны быть использованы в дальнейшем при изучении дисциплины «Биоинформатика для высокопроизводительного секвенирования»

# Содержание УЧЕБНОЙ ДИСЦИПЛИНЫ

**Тема 1. Введение, структуры данных и управляющие конструкции**

Вектора, листы, матрицы и дата-фреймы. Атрибуты переменных. Типы индексации. Объектно-ориентированый подход (система S3). Уловные выражения, циклы и функции семейства apply, пакет plyr. Функции, использование анонимные функций.

**Тема 2. Графика, статистические функции, биоинформатические пакеты**

Базовая графика, аннотация рисунков, сохранение рисунков в файл. Сглаживание и аппроксимация. Работа с цветом. Функции для статистических распределений. Т-тест, тест Вилкоксона, тест Фишера, линейные модели, анализ вариабельности. Обобщённые линейные модели. Поправка на множественное тестирование. Пакеты seqinr, ape, GenomicRanges.

# ОЦЕНИВАНИЕ

Каждое занятие состоит из лекции и практического занятия по результатам которого ставится оценка. На каждом занятии даётся домашнее задание, решение которого (код на R) должно быть сдано по электронной почте преподавателю не позже чем за сутки до следующего занятия. Домашние и практические задания оцениваются по бинарной шкале (0 или 1 балл). Курс включает две контрольных. Первая контрольная состоит из заданий базовому синтаксису и графике, вторая (экзаменационная) контрольная включает в себя задания на знания статистических методов. Каждая из двух контрольных включает в себя задачи различной сложности (6 или 10 баллов). Решение задачи сдаётся преподавателю непосредственно на контрольной. Задача считается решённой если код выполняет поставленную задачу, студент может объяснить что делает произвольно выбранная строка кода и может быстро внести небольшие изменения в код.

Итоговая оценка складывается из суммы оценок за практические, домашние и контрольные работы, при этом оценка за контрольные шкалируется так, чтобы две контрольных, с одной стороны, и все домашние и практические задания, с другой, дают равный вклад в итоговую оценку. Для приведения к десятичной шкале оценка делится на четыре и округляется в большую сторону.

# ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Оценочные средства для текущего контроля студента: Сгенирируйте случайное число, посчитайте его синус при помощи ряда Тейлора.

Оценочные средства для промежуточной аттестации: Данные в файле 1.txt являются числом попаданий в мишень за 10 мин стрельбы в зависимости от типа оружия, пола стреляющего и концентрации алкоголя у него в крови. При помощи GLM с распределение Пуассона определить какие факторы (комбинации факторов) влияют на вероятность попадания. Изобразить графически реальные данные и результаты предсказания модели.

1. **РЕСУРСЫ**
	1. **Основная литература**
	2. **Дополнительная литература**
	3. **Программное обеспечение**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
| 1. |  Microsoft Windows 7 Professional RUSMicrosoft Windows 10Microsoft Windows 8.1 Professional RUS | *Из внутренней сети университета (договор)* |
| 2. | Microsoft Office Professional Plus 2010 | *Из внутренней сети университета (договор)* |

* 1. **Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
|  | ***Профессиональные базы данных, информационно-справочные системы*** |
| 1. | Консультант Плюс | *Из внутренней сети университета (договор)* |
| 2. | Электронно-библиотечная система Юрайт  | URL: https://biblio-online.ru/ |
|  | ***Интернет-ресурсы (электронные образовательные ресурсы)*** |
| 1. | Открытое образование  | URL: https://openedu.ru/ |
| 2. | Advanced R by Hadley Wickham | URL: <http://adv-r.had.co.nz/> |

* 1. **Материально-техническое обеспечение дисциплины**

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);

мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.