


**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ»**

Утверждаю:  
Зав. кафедрой компьютерных  
интеллектуальных технологий  
проектирования

 М.И. Чижов  
«21» декабря 2021 г.

**УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ  
ПО ДИСЦИПЛИНЕ**

**«Язык R и базовая статистика»**

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность (профиль): Искусственный интеллект

Квалификация выпускника магистр

Нормативный период обучения 2 года / 2 года и 5 м.

Форма обучения очная / заочная

Год начала подготовки 2022

Составитель:  
ПЕТРОВ Р.В., Д.Ф.-М.Н., ДОЦЕНТ, ЗАВЕДУЮЩИЙ КАФЕДРОЙ  
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ (НГУ)  
ТЕЛИНА И.С., ДОЦЕНТ КАФ  
ФИЗИКИ ТВЕРДОГО ТЕЛА И МИКРОЭЛЕКТРОНИКИ (НГУ)  
Ершов Евгений Валентинович, д.т.н., профессор, директор  
института информационных технологий, зав. кафедрой МПО ЭВМ ЧГУ

г. Воронеж - 2021

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

1 Python и анализ данных [Электронный ресурс] / Маккинли Уэс, А. Слинкина ; Уэс Маккинли; пер. А. Слинкина. - Python и анализ данных ; 2024-10-28. - Саратов : Профобразование, 2019. - 482 с.

2 Анализ данных [Электронный ресурс] : Учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева ; Г. В. Шнарева, Ж. Г. Пономарева. - Анализ данных ; 2024-12-06. - Симферополь : Университет экономики и управления, 2019. - 129 с.

**Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)**

Ресурсы информационно-телекоммуникационной сети «Интернет»:

- <http://www.edu.ru/>

- Образовательный портал ВГТУ

Информационные справочные системы:

- <http://window.edu.ru>

- <https://wiki.cchgeu.ru/>

**Учебно-методические указания и рекомендации  
к изучению тем лекционных и практических занятий, самостоятельной  
работе студентов**

**Очная форма обучения**

№ п.п.	Наименование темы	Содержание раздела	Лекц.	Лаб. зан.	СРС	Всего час.
1	Знакомство с R Визуализация данных	Идеология R. Структуры данных (векторы, матрицы, массивы данных, таблицы данных, факторы, списки). Загрузка данных. Пакеты. Числовые и текстовые функции. Циклы и условия. Пользовательские функции.	2	-	12	20
2	Визуализация данных	Столбчатые диаграммы. Гистограммы. Диаграммы плотности распределений. Диаграммы размахов. Диаграммы рассеяния.	2	4	12	20
3	Разведочный анализ данных	Оценки центрального положения и вариабельности. Исследование распределений данных. Корреляция. Тест Стьюдента. Непараметрические тесты межгрупповых различий. Визуализация групповых различий.	2	4	12	18
4	Проверка статистических гипотез.	A/B-тестирование. Проверка статистических гипотез. Статистическая значимость и р-значение. Проверка на основе t-статистики. Дисперсионный анализ. Проверка на основе статистики хи-квадрат.	4	4	12	18
5	Регрессия и классификация в R	Простая линейная регрессия. Множественная линейная регрессия. Предсказания на основе регрессии. Оценка качества предсказания. Перекрестная проверка. Нелинейная регрессия. Навивный байесовский алгоритм. Логистическая регрессия. Предсказание значений в логистической регрессии. Оценка качества предсказания.	4	4	12	16
6	Статистическое машинное обучение.	Алгоритм k ближайших соседей. Деревозидные алгоритмы (дерево решений, случайный лес). Алгоритм бустинга. Переобучение. Подбор гиперпараметров алгоритм	2	4	12	16
<b>Итого</b>			<b>16</b>	<b>20</b>	<b>72</b>	<b>108</b>

### Заочная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего. час
1	Знакомство с R Визуализация данных	Идеология R. Структуры данных (векторы, матрицы, массивы данных, таблицы данных, факторы, списки). Загрузка данных. Пакеты. Числовые и текстовые функции. Циклы и условия. Пользовательские функции.	1	-	14	18
2	Визуализация данных	Столбчатые диаграммы. Гистограммы. Диаграммы плотности распределений. Диаграммы размахов. Диаграммы рассеяния.	1	1	14	18
3	Разведочный анализ данных	Оценки центрального положения и вариативности. Исследование распределений данных. Корреляция. Тест Стьюдента. Непараметрические тесты межгрупповых различий. Визуализация групповых различий.	1	1	14	18
4	Проверка статистических гипотез.	A/B-тестирование. Проверка статистических гипотез. Статистическая значимость и р-значение. Проверка на основе t-статистики. Дисперсионный анализ. Проверка на основе статистики хи-квадрат.	2	2	14	18
5	Регрессия и классификация в R	Простая линейная регрессия. Множественная линейная регрессия. Предсказания на основе регрессии. Оценка качества предсказания. Перекрестная проверка. Нелинейная регрессия. Наивный байесовский алгоритм. Логистическая регрессия. Предсказание значений в логистической регрессии. Оценка качества предсказания.	2	2	16	16
6	Статистическое машинное обучение.	Алгоритм k ближайших соседей. Древозидные алгоритмы (дерево решений, случайный лес). Алгоритм бустинга. Переобучение. Подбор гиперпараметров алгоритм	1	2	16	16
<b>Итого</b>			<b>8</b>	<b>8</b>	<b>88</b>	<b>104</b>

### Темы лабораторных работ

**Лабораторная работа № 1.** Установка RStudio. Работа с основными типами данных. Создание датафреймов и загрузка внешних данных. Загрузка пакетов и написание функций. Построение графиков с использованием пакета ggplot2

**Лабораторная работа № 2.** Расчет описательных статистик и визуализация распределений

**Лабораторная работа № 3.** Тестирование гипотез

**Лабораторная работа № 4.** Предсказание на основе уравнения линейной регрессии

**Лабораторная работа № 5.** Предсказание значений в логистической регрессии. Использование алгоритма XGBoost для предсказания значений

## Контрольные вопросы к лабораторным работам

1. Назовите основные типы данных языка R.
2. Какими методами можно создать датафрейм?
3. Приведите алгоритм построения графиков.
4. Способы визуализации распределений.
5. Методы расчета описательных статистик.
6. Методика тестирования гипотез.
7. Опишите алгоритм предсказания на основе уравнения линейной регрессии.
8. Опишите алгоритм для предсказания значений в логистической регрессии.

## Средства контроля качества обучения

### Вопросы к зачету

- 1 Идеология R. Структуры данных (векторы, матрицы, массивы данных, таблицы данных, факторы, списки).
- 2 Загрузка данных. Пакеты. Числовые и текстовые функции. Циклы и условия.
- 3 Пользовательские функции
- 4 Визуализация данных. Диаграммы. Гистограммы.
- 5 Визуализация данных. Диаграммы плотности распределений. Диаграммы размахов. Диаграммы рассеяния
- 6 Оценки центрального положения и вариабельности.
- 7 Исследование распределений данных.
- 8 Корреляция. Тест Стьюдента.
- 9 Непараметрические тесты межгрупповых различий.
- 10 Визуализация групповых различий
- 11 А/В-тестирование. Проверка статистических гипотез.
- 12 Статистическая значимость и р-значение.
- 13 Проверка на основе t-статистики.
- 14 Дисперсионный анализ.
- 15 Проверка на основе статистики хи-квадрат
- 16 Простая линейная регрессия.
- 17 Множественная линейная

регрессия.

18 Предсказания на основе регрессии.

19 Оценка качества предсказания.

Перекрестная проверка.

20 Нелинейная регрессия

21 Наивный байесовский алгоритм.

22 Логистическая регрессия.

Предсказание значений в логистической регрессии. Оценка качества предсказания

23 Алгоритм k ближайших соседей.

24 Древоподобные алгоритмы (дерево решений, случайный лес).

25 Алгоритм бустинга.

26 Переобучение. Подбор гиперпараметров алгоритма