

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Ректор Белорусского
государственного университета

А.Д.Король



СТАТИСТИЧЕСКИЙ АНАЛИЗ И ЯЗЫК R

Учебная программа учреждения образования по учебной дисциплине для
специальности:

1-31 03 08 Математика и информационные технологии (по направлениям)

Направления специальности:

1-31 03 08-01 Математика и информационные технологии (Веб-
программирование и интернет-технологии)

2024 г.

Учебная программа составлена на основе ОСВО 1-31 03 08-2021, учебных планов: №G31-1-011/уч. от 25.05.2021, №G31-1-003/уч.ин. от 31.05.2021, №G31-1-220/уч. от 22.03.2022, №G31-1-225 уч.ин. от 27.05.2022.

СОСТАВИТЕЛЬ:

Дубров Б.М., доцент кафедры веб-технологий и компьютерного моделирования механико-математического факультета Белорусского государственного университета.

РЕЦЕНЗЕНТ:

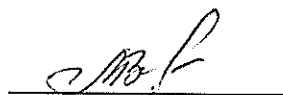
Куликов С.С., доцент кафедры ПОИТ Белорусского государственного университета информатики и радиотехники Республики Беларусь, кандидат технических наук, доцент.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой веб-технологий и компьютерного моделирования БГУ
(протокол № 9 от 24.05.2024)

Научно-методическим советом БГУ
(протокол № 9 от 28.06.2024)

Заведующий кафедрой
Кандидат физ.-мат. наук, доцент



М.В. Игнатенко

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины создание у студентов компетенций в области использования языка программирования R для статистического анализа данных

Задачи учебной дисциплины:

1. Знакомство с основными моделями статистического анализа данных.
2. Получение практических навыков использования языка программирования R.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием. Дисциплина «Статистический анализ и язык R» относится к **дисциплинам специализаций** компонента учреждения высшего образования.

Связи с другими учебными дисциплинами. Изучение дисциплины «Статистический анализ и язык R» базируется на знаниях, полученных при изучении дисциплин «Методы программирования», «Теория вероятностей и математическая статистика».

Требования к компетенциям

Освоение учебной дисциплины «Статистический анализ и язык R» должно обеспечить формирование следующих универсальных, базовых профессиональных и специализированных компетенций:

универсальные компетенции:

- Владеть основами исследовательской деятельности, осуществлять поиск, анализ и синтез информации;
- Решать стандартные задачи профессиональной деятельности на основе применения информационно-коммуникационных технологий;
- Быть способным к саморазвитию и совершенствованию в профессиональной деятельности;
- Проявлять инициативу и адаптироваться к изменениям в профессиональной деятельности;

базовые профессиональные компетенции:

- Применять теоретические знания и навыки в самостоятельной исследовательской деятельности;
- Применять современные технологии и базовые конструкции языков программирования для реализации алгоритмических прикладных задач и разработки веб-проектов;

специализированные компетенции:

- Выполнять проектирование, разработку, тестирование, и маркетинг информационных решений в сети Интернет с учетом их последующего масштабирования и обработки возникающих больших объемов данных.

В результате освоения учебной дисциплины студент должен:

знать:

– модели статистического анализа данных и условия их применения; основные типы данных и управляющие конструкции языка программирования R;

уметь:

– формулировать статистические гипотезы; выбирать подходящие инструменты для анализа данных; грамотно формулировать полученные результаты;

владеть:

– средствами языка программирования R для статистического анализа данных.

Структура учебной дисциплины

Дисциплина изучается в 7. Всего на изучение учебной дисциплины «Статистический анализ и язык R» отведено:

— в очной форме получения высшего образования: 90 часов, в том числе 36 аудиторных часа, из них: лекции — 18 часов, лабораторные занятия — 16 часов, управляемая самостоятельная работа — 2 часа.

Трудоемкость учебной дисциплины составляет 3 зачетных единиц.

Форма промежуточной аттестации — зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Введение в статистический анализ

Статистика в повседневной жизни. Краткое введение в дизайн исследования. Основы статистического измерения. Надежность и валидность измерения. Предикторы. Распространенные распределения: нормальное, биномиальное, t-Стьюдента, χ^2 (хи-квадрат), F-распределение Фишера.

Тема 2. Введение в язык R

Знакомство с R и Rstudio. Ввод команд в консоли. Простые вычисления. Использование функций. Введение в переменные. Числовые, символьные и логические данные. Хранение нескольких значений в виде вектора. Установка и загрузка пакетов. Сложные структуры данных: факторы, фреймы данных, списки и формулы.

Тема 3. Описательная статистика

Среднее, медиана и мода. Вариационный размах, межквартильный размах, дисперсия и стандартное квадратическое отклонение. Асимметрия и эксцесс. Стандартные оценки. Корреляции. Инструменты для вычисления этих свойств в R. Отсутствующие данные.

Тема 4. Использование графических средств

Рисование графиков в R. Гистограммы. Диаграммы стебля и листьев. Коробчатая диаграмма. Столбчатая диаграмма.

Тема 5. Продвинутое возможности языка R

Преобразование переменной. Подмножество векторов и фреймов данных. Сортировка, транспонирование и слияние данных. Изменение формы фрейма данных. Основы обработки текста. Чтение необычных файлов данных. Основы приведения переменных. Основы программирования. Скрипты. Циклы. Условные операторы. Написание функций. Неявные циклы.

Тема 6. Проверка статистических гипотез

Выборка из популяций. Оценка средних значений и стандартных отклонений популяций. Распределения выборки. Центральная предельная теорема. Доверительные интервалы. Нулевые и альтернативные гипотезы. Ошибки типа I и типа II. Распределения выборки для статистики теста. Проверка гипотез как принятие решений. р-значения. Отчет о результатах теста. Размер и мощность эффекта.

Тема 7. Сравнение средних и однофакторный дисперсионный анализ

Одновыборочный z-тест. Одновыборочный t-тест. Независимый выборочный t-тест Стьюдента. Парный выборочный t-тест. Размер эффекта и d Коэна. Проверка предположения о нормальности. Тесты Вилкоксона для ненормальных данных. Введение в однофакторный дисперсионный анализ (ANOVA). Предположения однофакторного ANOVA. Проверка однородности дисперсии с использованием тестов Левена. Избегание предположения об однородности дисперсии. Проверка и избегание предположения о нормальности. Связь между ANOVA и t-тестами.

Тема 8. Линейная регрессия и многофакторный дисперсионный анализ

Введение в линейную регрессию. Метод наименьших квадратов. Измерение соответствия модели регрессии. Тесты гипотез для моделей регрессии. Стандартизированный коэффициент регрессии. Предположения моделей регрессии. Базовая диагностика регрессии. Методы выбора модели для регрессии.

Многофакторный дисперсионный анализ (MANOVA). Модель взаимодействия. Размеры эффекта, доверительные интервалы для эффектов. Проверка предположений. F-тесты как выбор модели. Интерпретация MANOVA как линейной модели.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Очная форма получения высшего образования с применением дистанционных образовательных технологий (ДОТ)

Номер темы	Название темы	Количество аудиторных часов					Количество часов УСР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия			
1	2	3	4	5	6	7	8	
1	Введение в статистический анализ	2						
2	Введение в язык R	2			2			Письменный отчет с устной защитой лабораторной работы
3	Описательная статистика	2			2			Контрольная работа
4	Использование графических средств	2			2			Письменный отчет с устной защитой лабораторной работы
5	Продвинутые возможности языка R	2			2			Контрольная работа
6	Проверка статистических гипотез	2			2			Опрос, письменный отчет с устной защитой лабораторной работы
7	Сравнение средних и однофакторный дисперсионный анализ	2			2	2		Контрольная работа
8	Линейная регрессия и многофакторный дисперсионный анализ	4			4			Опрос, контрольная работа

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная литература

1. Васильев А. А. Теория вероятностей и математическая статистика. М.: Юрайт, 2023. 225 с.
2. Кельберт М. Я., Сухов Ю. М. Вероятность и статистика в примерах и задачах. Том I. Основные понятия теории вероятностей и математической статистики. Часть 2. Статистические модели. М.: МЦНМО, 2024. 352 с.
3. Long JD., Teetor P. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics. O'Reilly Media; 2nd edition. 2019. 598p.

Дополнительная литература

1. Navarro, D. Learning Statistics with R, Онлайн ресурс: <https://learningstatisticswithr.com/>.
2. Кремер Н.Ш. Теория вероятностей и математическая статистика. М.: Юнити, 2004. 573 с.

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Объектом диагностики компетенций студентов являются знания, умения, полученные ими в результате изучения учебной дисциплины. Выявление учебных достижений студентов осуществляется с помощью мероприятий текущего контроля и промежуточной аттестации.

Для текущего контроля качества усвоения знаний по дисциплине могут использоваться следующие средства текущего контроля:

- устный опрос;
- письменный отчет с устной защитой лабораторной работы;
- контрольная работа.

Формой промежуточной аттестации по дисциплине «Статистический анализ и язык R» учебным планом предусмотрен зачет.

Примерная тематика лабораторных работ

1. Синтаксис, типы данных, управляющие конструкции языка R.
2. Визуализация описательной статистики с помощью диаграмм.
3. Управление фреймами данных: фильтрация, сортировка, слияние, транспонирование.
4. Вычисление мощности и размера эффекта при проверке статистических гипотез.
5. Проверка предположения нормальности распределения данных.
6. Независимый и парный t-тест Стьюдента на языке R.

7. Однофакторный дисперсионный анализ.
8. Линейная регрессия.
9. Многофакторный дисперсионный анализ.
10. Модели взаимодействия независимых переменных.

Примерный перечень заданий для управляемой самостоятельной работы студентов

Тема 7. Сравнение средних и однофакторный дисперсионный анализ

Задание. Проверить предположения и выполнить сравнение средних для двух заданных выборок данных. Сгенерировать отчет с использованием языка R.

Форма контроля - Контрольная работа

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используется **практико-ориентированный подход**, который предполагает:

освоение содержание образования через решения практических задач; приобретение навыков эффективного выполнения разных видов профессиональной деятельности; ориентацию на генерирование идей, реализацию групповых студенческих проектов.

Методические рекомендации по организации самостоятельной работы

При изучении учебной дисциплины рекомендуется использовать следующие формы самостоятельной работы:

- поиск (подбор) и обзор литературы и электронных источников по индивидуально заданной проблеме учебной дисциплины;
- выполнение домашнего задания;
- научно-исследовательские работы;
- анализ фактических материалов по заданной теме, проведение расчетов, составление схем и моделей на основе фактических материалов;
- подготовка к участию в конференциях и конкурсах.

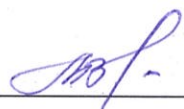
Примерный перечень вопросов к зачету

1. Синтаксис, типы данных языка R.
2. Управляющие конструкции языка R.
3. Управление фреймами данных: фильтрация, сортировка, слияние, транспонирование.
4. Описательные статистики: среднее, медиана, квартили и перцентили.
5. Визуализация описательной статистики с помощью диаграмм в языке R.
6. Принципы проверки статистических гипотез.
7. Построение доверительных интервалов. Значение p .
8. Ошибки первого и второго рода при проверке статистических гипотез.
9. Распределение Стюдента.
10. Хи-квадрат распределение.
11. F-распределение Фишера.
12. Проверка предположения нормальности распределения данных.
13. Независимый t -тест Стюдента.
14. Парный t -тест Стюдента.
15. Хи-квадрат тест согласия.
16. Хи-квадрат тест независимости.
17. Однофакторный дисперсионный анализ.
18. Двухфакторных дисперсионный анализ.
19. Линейная регрессия. Проверка значимости линейной модели.
20. Построение интервальных оценок для коэффициентов линейной регрессии.
21. Дисперсионный анализ на языке линейной модели.
22. Взаимодействие независимых переменных в линейной модели.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Язык программирования Python	Веб-технологий и компьютерного моделирования	Изменений не требуется	Протокол № 9 от 24.05.2024

Заведующий кафедрой
Кандидат физ.-мат. наук, доцент



М.В. Игнатенко

24. мая 2024 г.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ
ПО ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 202_ г.)
(название кафедры)

Заведующий кафедрой

(ученая степень, ученое звание)

(И.О.Фамилия)

УТВЕРЖДАЮ
Декан факультета

(ученая степень, ученое звание)

(И.О.Фамилия)