

Белорусский государственный университет

УТВЕРЖДАЮ

Проректор по учебной работе и
образовательным инновациям

О.И. Чуприс

2018 г.

Регистрационный № УД-5918 /уч.



**ОСНОВЫ КОМПЬЮТЕРНОГО АНАЛИЗА ДАННЫХ С
ИСПОЛЬЗОВАНИЕМ ЯЗЫКА R**

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности второй степени высшего
образования (магистратуры) с углубленной подготовкой специалиста:**

**1-31 81 09 Алгоритмы и системы обработки больших объемов
информации**

2018 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 81 09-2014 и учебного плана G31-254/уч. от 26.05.2017.

СОСТАВИТЕЛЬ:

И.А. Бодягин, заведующий кафедрой математического моделирования и анализа данных факультета прикладной математики и информатики Белорусского государственного университета, кандидат физико-математических наук, доцент.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой математического моделирования и анализа данных Белорусского государственного университета (протокол № 13 от 29.03.2018);

Научно-методическим Советом Белорусского государственного университета (протокол № 7 от 13 июля 2018 г.).

Зав. кафедрой ИМАД *И.А. Бодягин*
И.А. Бодягин

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цель преподавания учебной дисциплины – ознакомление студентов II ступени (магистрантов) с основными возможностями и синтаксисом скриптового языка программирования R, а также с методами решения основных прикладных задач статистического анализа данных.

В рамках поставленной цели **задачи** учебной дисциплины состоят в следующем:

- 1) изучение синтаксиса и языка программирования R;
- 2) формирование практических умений и навыков работы в графической среде разработки RStudio.
- 3) приобретение практических навыков для решения прикладных задач статистического анализа данных

Учебная дисциплина «Основы компьютерного анализа данных с использованием языка R» относится к циклу дисциплин специальной подготовки компоненты учреждения высшего образования.

Учебная программа составлена с учетом межпредметных связей с учебными дисциплинами. Так, основой для изучения дисциплины «Основы компьютерного анализа данных с использованием языка R» является дисциплина первой ступени «Теория вероятностей и математическая статистика». Знания, полученные в результате изучения дисциплины, будут использованы при изучении дисциплины II ступени высшего образования «Методы машинного обучения», а также способствовать успешному прохождению производственной практики по специальности и подготовки магистерской диссертации.

В результате освоения учебной дисциплины студент магистратуры должен:

знать:

- основные возможности языка программирования R;
- методы решения основных задач первичного статистического анализа данных;

уметь:

- создавать сложные скрипты (программы) на языке R, содержащие циклы;
- работать с данными разных размерностей и форматов,
- выполнять работу с графиками;
- готовить отчеты в среде языка программирования R;

владеть:

- методами решения типовых задач предварительного статистического анализа;
- навыками программирования с помощью языка R.

Освоение учебной дисциплины «Основы компьютерного анализа данных с использованием языка R» должно обеспечить формирование

следующих академических, социально-личностных и профессиональных компетенций:

академические компетенции:

АК-2. Методологические знания и исследовательские умения, обеспечивающие решение прикладных задач и инновационной деятельности.

социально-личностные компетенции:

СЛК-1. Учитывать социальные и нравственно-этические нормы в социально-профессиональной деятельности.

СЛК-2. Быть способным к сотрудничеству и работе в команде.

СЛК-3. Владеть коммуникативными способностями для работы в междисциплинарной и международной среде.

СЛК-4. Совершенствовать и развивать свой интеллектуальный и общекультурный уровень, добиваться нравственного и физического совершенствования своей личности.

СЛК-6. Логично, аргументированно и ясно строить устную и письменную речь, использовать навыки публичной речи, ведения дискуссии и полемики.

СЛК-7. Проявлять инициативу и креативность, в том числе в нестандартных ситуациях.

профессиональные компетенции:

ПК-1. Квалифицированно использовать современные достижения по разработке и анализу алгоритмов и современные информационные технологии.

ПК-3. Разрабатывать эффективные численные алгоритмы и интегрировать их в компьютерные системы.

ПК-4. Обосновывать выбор методов и инструментов для решения прикладных задач.

Структура содержания учебной дисциплины включает такие дидактические единицы, как темы (разделы), в соответствии с которыми разрабатываются и реализуются соответствующие лекционные и семинарские занятия. Примерная тематика занятий приведена в информационно-методической части.

Дисциплина изучается в 1 семестре (II ступень). Всего на освоение учебной дисциплины «Основы компьютерного анализа данных с использованием языка R» отведено 120 часов, в том числе 40 аудиторных часов, из них: лекции – 20 часов, лабораторные занятия – 20.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма текущей аттестации – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел I. Введение

Тема 1.1. Общая информация о языке R. История создания языка R. Сравнение возможностей языка R и других пакетов, предназначенных для статистического анализа данных.

Тема 1.2. Основы работы с языком R. Консоль. Настройка рабочей среды. Получение справки. Работа в RStudio. Создание, хранение и удаление объектов из памяти.

Раздел II. Основы языка программирования R

Тема 2.1. Типы данных в R. Основные типы данных (числовые, логические, символьные), векторы, факторы, матрицы. Работа с векторами и матрицами. Списки и таблицы данных. Создание списков и таблиц. Доступ к элементам списка или таблицы. Внесение новых данных в таблицу.

Тема 2.2. Обработка данных. Векторные вычисления, принцип циклического повторения ("re-cycling"). Основные операции. Операции над матрицами. Сортировка. Работа с пропущенными значениями. Применение функций вида apply.

Тема 2.3. Условные операторы выбора и операторы циклов. Условный оператор выбора if else. Условный оператор множественного выбора switch. Операторы циклов for, while, repeat.

Тема 2.4. Функции. Создание новых функций, определенных пользователем. Область видимости переменных. Именованные аргументы, значения по-умолчанию, произвольное количество аргументов. Определение новых бинарных операторов. Функция в качестве аргумента функции. Функция, возвращающая функцию. Рекурсия.

Тема 2.5. Обработка исключительных ситуаций. Настройки ядра языка R для работы с исключениями. Диагностические сообщения. Предупреждающие сообщения. Фатальные ошибки. Функция tryCatch.

Тема 2.6. Импорт и экспорт данных. Редактор данных. Доступ к внешним данным. Чтение данных с клавиатуры. Чтение данных из внешнего файла, страницы в сети Интернет. Сохранение данных в файл.

Тема 2.7. Подготовка отчетов в языке R. Подготовка отчета в среде LaTeX, содержащего скрипты языка R. Пакет Sweave. Работа с документами в разных кодировках. Подготовка отчета в среде RStudio с помощью пакета RMarkdown.

Раздел III. Применение языка R для статистического анализа данных

Тема 3.1. Простейший статистический анализ. Вычисление базовых описательных статистик. Проверка гипотез о принадлежности выборки заданному распределению.

Тема 3.2. Работа с графиками. Понятие графического устройства в R. Построение простейших графиков. Построение классические графиков (гистограмм, ящиков с усами, графиков рассеяния и пр.). Точная настройка атрибутов и свойств графиков.

Тема 3.3. Выборочные тесты. t-критерии Стьюдента (одновыборочный, двухвыборочный для независимых выборок, двухвыборочный для зависимых выборок и пр.), t-критерии Вилкоксона.

Тема 3.4. Генерация псевдослучайных чисел. Генерация базовой случайной величины. Генерация дискретных случайных величин. Методы генерации непрерывных случайных величин. Генерация Гауссовский случайных величин. Метод Монте-Карло.

Тема 3.5. Применение языка R для решения задачи кластерного и дискриминантного анализа. Формулировка задач кластерного и дискриминантного анализ. Методы их решения в языке программирования R.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Но мер раз дел а, тем ы	Название раздела, темы	Количество аудиторных часов			Форма контроля знаний
		Лек ции	Семи нарск ие Занят ия	Лабо ратор ные занят ия	
1	Введение	2	0	0	
1.1	Общая информация о языке R.	1	0		Устный опрос
1.2	Основы работы с языком R.	1			Устный опрос
2	Основные типы статистических графиков	12		16	
2.1	Типы данных в R.	2			Устный опрос
	<i>Лабораторная работа 1. Работа с объектами разных типов данных</i>			4	Защита лабораторной работы
2.2	Обработка данных	2			Устный опрос
2.3	Условные операторы выбора и операторы циклов	2			Устный опрос
	<i>Лабораторная работа 2. Решение математических задач средствами языка R</i>			4	Защита лабораторной работы
2.4	Функции	2			Контрольная работа № 1
	<i>Лабораторная работа 3. Решение алгоритмических задач</i>			4	Защита лабораторной работы
2.5	Обработка исключительных ситуаций	1			Отчет по заданию с устной защитой
2.6	Импорт и экспорт данных.	1			Устный опрос
2.7	Подготовка отчетов в языке R.	2			Коллоквиум
	<i>Лабораторная работа 4. Подготовка отчетов средствами языка R</i>			4	Защита лабораторной работы
3	Применение языка R для статистического анализа данных	6		4	

3.1	Простейший статистический анализ	1			Устный опрос
3.2	Работа с графиками	1			Устный опрос
3.3	Выборочные тесты	1			Устный опрос
3.4	Генерация псевдослучайных чисел	1			Защита подготовленного реферата
	<i>Лабораторная работа 5 Разведывательный статистический анализ</i>			4	Защита лабораторной работы
3.5	Применение языка R для решения задачи кластерного и дискриминантного анализа.	2			Контрольная работа №2.
ИТОГО		20	0	20	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. М.: ДМК-пресс, 2015. – 496 с.
2. Owen Jones, Robert Maillardet, Andrew Robinson. Scientific Programming and Simulation Using R. – London, New York: CRC-Press, 2009. – 455 p.
3. Andy Field, Jeremy Miles, Zoe Field. Discovering statistics using R. – London: SAGE Publications Ltd, 2012. – 1193 p.
4. Peter Dalgaard. Introductory Statistics with R. – New York: Springer, 2008. – 370 p.

Перечень дополнительной литературы

1. P. Filzmoser. Linear and nonlinear methods for regression and classification and applications in R. – Vienna University of Technology. CS. №3, 2008. – 1-52 pp.
2. D. Navarro. Learning statistics with R: A tutorial for psychology students and other beginners., 2013. – 542 p.

Рекомендуемая тематика контрольных работ

- 1) Контрольная работа №1. *Применение скриптового языка R для решения классических задач программирования*
- 2) Контрольная работа №2. *Применение языка R для решения прикладных задач статистического анализа данных.*

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов магистратуры по учебной дисциплине следует использовать современные информационные технологии: разместить в сетевом доступе комплекс учебных и учебно-методических материалов (учебно-программные материалы, ссылки на учебные издания для теоретического изучения дисциплины, методические указания к лабораторным занятиям, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к зачету, задания, тесты, вопросы для самоконтроля, тематика рефератов и др., список рекомендуемой литературы, информационных

ресурсов и др.). Эффективность самостоятельной работы студентов магистратуры проверяется в ходе текущего и итогового контроля знаний. Для общей оценки качества усвоения студентами магистратуры учебного материала рекомендуется использование рейтинговой системы.

Перечень рекомендуемых средств диагностики

Для текущего контроля качества усвоения знаний студентами магистратуры используется следующий диагностический инструментарий:

1. Устная форма: устные опросы; защиты отчетов по домашним заданиям, при выполнении студентами магистратуры лабораторных работ; проведение коллоквиума; защита подготовленного студентом магистратуры реферата (рефераты используются для обобщения и систематизации учебного материала; в процессе подготовки реферата студент магистратуры мобилизует и актуализирует имеющиеся умения, приобретает самостоятельно новые знания, необходимые для раскрытия темы, сопоставляя разные позиции и точки зрения).

2. Письменная форма: письменные контрольные работы по отдельным темам учебной дисциплины.

Методика формирования итоговой оценки

Формой текущей аттестации по учебной дисциплине «Основы компьютерного анализа данных с использованием языка R» учебным планом предусмотрен зачет.

Рекомендуется использовать рейтинговую оценку знаний студента магистратуры, дающую возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине. Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний в рейтинговую оценку:

- подготовка реферата – 15 %;
- работа на лабораторных занятиях – 35 %;
- контрольные работы – 25 %;
- коллоквиум – 15 %.

Итоговая оценка формируется на основе:

- 1) Правил проведения аттестации студентов (Постановление Министерства образования Республики Беларусь № 53 от 29 мая 2012г.);
- 2) Положение о рейтинговой системе оценки знаний по дисциплине в БГУ (Приказ ректора БГУ от 18.08.2015 № 382-ОД);
- 3) Критериев оценки знаний студентов (письмо Министерства образования от 22.12.2003).

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

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)
Методы визуализации в анализе больших данных	Дискретной математики и алгоритмики	нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 29.03.2018 г.

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

на 2018/2019 учебный год

№ п/п	Дополнения и изменения	Основание
1	Добавить в перечень основной литература: 1. Котов В.В., Себастьянова Е.П., Толстиков А.А. Нормы и структура данных: учебное пособие. - Минск, БГУ, 2012	Учебное пособие содержит базовый материал, необходимый для изучения дисциплины «Основы компьютерного анализа данных с использованием Excel».

Учебная программа пересмотрена и одобрена на заседании кафедры
М.М.АД (протокол № 8 от 04.12 2018 г.)

Заведующий кафедрой

М.М.АД

Т.И.А.

И.А.Богачев

УТВЕРЖДАЮ

И.И. Декан факультета

И.И.

И.И.

Т.В.Васильев