

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 15.07.2024 14:54:55

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51517ff6d591

Автономная некоммерческая образовательная организация высшего образования
«Европейский университет в Санкт-Петербурге»

УТВЕРЖДАЮ

Ректор / В.В. Волков

« 28 февраля » 2024 г.

Протокол Ученого Совета

№ 2 от 28 февраля 2024 г.



Рабочая программа дисциплины
«Программирование в R»

дополнительная профессиональная программа
«Прикладной анализ данных»

вид программы
программа профессиональной переподготовки

язык обучения – русский
форма обучения – очная

Санкт-Петербург

Авторы:

Тушканова О.Н., к.т.н., доцент факультета социологии АНООВО «ЕУСПб»

Рабочая программа дисциплины «Программирование в R», входящая в состав дополнительной профессиональной программы профессиональной переподготовки «Прикладной анализ данных» утверждена на заседании Ученого совета.

Содержание

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ. ОБЪЕМ ДИСЦИПЛИНЫ.....	4
2. ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ.....	4
3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ.....	5
4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ	7
5. ОЦЕНОЧНЫЕ СРЕДСТВА.....	8
6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	10
7. ПРОГРАММНОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММЫ.....	11

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ. ОБЪЕМ ДИСЦИПЛИНЫ

Цель обучения: освоения дисциплины «Программирование в R» — научить слушателей основам работы с языком программирования R, показать разные парадигмы программирования в R.

Задачи обучения:

- показать слушателям преимущества языка программирования R и базовую специфику работы с ним;
- научить работать с тремя основными парадигмами программирования в R: base-R, dplyr, data.table;
- обучить особенностям работы с циклами в R и решению задач через векторизацию.

Изучение данной дисциплины способствует формированию профессиональных навыков по разработке теоретических моделей исследуемых процессов, явлений и объектов, относящихся к сфере профессиональной деятельности, их исследованию, оценке и интерпретации полученных результатов.

Отличительной особенностью реализуемого подхода к преподаванию дисциплины является разнообразных практических иллюстраций основных теоретических положений.

Общая трудоемкость дисциплины составляет 62 часа.

2. ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ

В результате освоения дисциплины слушатель должен приобрести следующие знания и умения, необходимые для качественного изменения профессиональных компетенций:

слушатель должен знать:

- методы сбора и обработки данных;
- базовую специфику работы в R;
- особенности применения различных теоретико-методологических концепций с использованием технологий прикладного анализа данных;
- общие правила ведения научных дискуссий.

слушатель должен уметь:

- синтезировать новое профессиональное знание на базе применения знаний и аналитических навыков с использованием технологий прикладного анализа данных;
- использовать полученные знания и умения в профессиональной деятельности;
- уметь использовать преимущества языка программирования R.

слушатель должен владеть:

- методами сбора и обработки данных, в том числе с использованием технологий прикладного анализа данных;
- углублёнными теоретическими знаниями и практическими навыками программирования в R.

В результате изучения дисциплины «Программирование в R и основы статистики» слушатель приобретает следующие профессиональные компетенции (Таблица 1):

Планируемые результаты обучения по дисциплине

Таблица 1

Код и название компетенции	Содержание компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапам формирования компетенций
ПК-5	способен применить анализ данных к научным и общественным задачам	Знать: З (ПК-5) – основы анализа данных
		Уметь: У (ПК-5) – использовать методики анализа данных применительно к различным типам данных

Код и название компетенции	Содержание компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапам формирования компетенций
		Владеть: В (ПК-5) – навыками анализа данных с учетом специфики научных и общественных задач

3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Содержание дисциплины соотносится с планируемыми результатами обучения по дисциплине через задачи, формируемые компетенции и их компоненты (знания, умения, навыки – далее ЗУВ) в соответствии с таблицей 2.

Таблица 2

Содержание дисциплины

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Коды ЗУВ (в соответствии с табл. 1)
1	Введение	R как язык программирования. Функциональное программирование vs. ООП. Знакомство с R-Studio. Основные типы данных в R (num, int, factor, char, bool, date&time). Основные типы структур данных (vector, data.frame, list). Основные типы библиотек в R и их структура. Объяснение принципов векторизации. Базовые функции в R. Как загружать данные разных типов из разных источников. Выполнение в R простейших операций, которые слушатели уже умеют на языке Python (ifelse, gsub, grepl и проч.)	ПК-5	З (ПК-5) У (ПК-5) В (ПК-5)
2	Основы R	Как работать с типом данных “data.frame”? Обращение к переменным и строкам, индексация. Фильтрация, выборка переменных. Преобразование данных, long2wide/wide2long: merge, rbind, cbind, aggregate, dcast. Практика объединений и преобразования датафреймов. Повторение пройденного на предыдущем занятии вместе с новым материалом (фильтрация и преобразование переменных с помощью функций ifelse, gsub и проч.).	ПК-5	З (ПК-5) У (ПК-5) В (ПК-5)
3	Библиотека Dplyr	Объяснение общей логики работы в этой библиотеки. Повторение старого материала на dplyr. Основы визуализации данных в ggplot2	ПК-5	З (ПК-5) У (ПК-5) В (ПК-5)
4	Data.table	Преимущества data.table. Основные принципы работы и специфические операторы data.table. Повторение предыдущего материала на базе этой библиотеки.	ПК-5	З (ПК-5) У (ПК-5) В (ПК-5)
5	Циклы, специфика apply family	Повторение материала по data.table. Объяснения циклов в R (предполагается, что слушатели уже знают о них после курса по Python). Работа с разными типами apply family. Введение в работу с листами в R (иллюстрация работы типа apply family - lapply)	ПК-5	З (ПК-5) У (ПК-5) В (ПК-5)
6	Методы кластеризации и кластеризация в R	Основные методы кластеризации и их применение в R: K-Means, Mean-Shift, DBSCAN.	ПК-5	З (ПК-5) У (ПК-5) В (ПК-5)

Структура дисциплины

Таблица 3.

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины, час.				Форма текущего контроля успеваемости, промежуточной аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР ¹	
				Л	СЗ ²		
1.	Введение	R как язык программирования. Функциональное программирование vs. ООП. Знакомство с R-Studio. Основные типы данных в R (num, int, factor, char, bool, date&time). Основные типы структур данных (vector, data.frame, list). Основные типы библиотек в R и их структура. Объяснение принципов векторизации. Базовые функции в R. Как загружать данные разных типов из разных источников. Выполнение в R простейших операций, которые слушатели уже умеют на языке Python (ifelse, gsub, grepl и проч.)	10	3	2	5	Практические задания
2.	Основы R	Как работать с типом данных "data.frame"? Обращение к переменным и строкам, индексация. Фильтрация, выборка переменных. Преобразование данных, long2wide/wide2long: merge, rbind, cbind, aggregate, dcast. Практика объединений и преобразования датафреймов. Повторение пройденного на предыдущем занятии вместе с новым материалом (фильтрация и преобразование переменных с помощью функций ifelse, gsub и проч.).	10	2	3	5	
3.	Библиотека Dplyr	Объяснение общей логики работы в этой библиотеке. Повторение старого материала на dplyr. Основы визуализации данных в ggplot2	10	3	2	5	
4.	Data.table	Преимущества data.table. Основные принципы работы и специфические операторы data.table. Повторение предыдущего материала на базе этой библиотеки.	10	2	3	5	
5.	Циклы, специфика apply family	Повторение материала по data.table. Объяснения циклов в R (предполагается, что слушатели уже знают о них после курса по Python). Работа с разными типами apply family. Введение в работу с листами в R (иллюстрация работы типа apply family - lapply)	10	2	2	6	

¹ Самостоятельная работа, включает в себя часы на промежуточный контроль

² Могут включать в себя: лабораторные работы, круглые столы, мастер-классы, мастерские, деловые игры, ролевые игры, тренинги, семинары по обмену опытом, выездные занятия, консультации

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины, час.			Форма текущего контроля успеваемости, промежуточной аттестации	
			Всего	Аудиторная работа по видам учебных занятий			СР ¹
				Л	СЗ ²		
6.	Методы кластеризации и кластеризация в R	Основные методы кластеризации и их применение в R: K-Means, Mean-Shift, DBSCAN.	10	2	2	6	
7.	Промежуточная аттестация	Проект	2	-	-	2	
Всего:			62	14	14	34	

4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Общие положения.

Знания и навыки, полученные в результате лекций и семинарских занятий, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения исследовательской литературы (из списков основной, дополнительной), статей по проблематики занятия и их анализа.

Самостоятельная работа обучающегося представляет самостоятельное изучение дополнительных материалов, Интернет-ресурсов и пр. Подготовка к семинарским занятиям, создание докладов, проектов и презентаций также является важной формой работы обучающихся. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя. Вопросы и замечания, возникшие в ходе самостоятельного внеаудиторного чтения рекомендованной литературы, обсуждаются с преподавателем и другими обучающимися. Выносятся на обсуждение, как правило, актуальные проблемы и предлагается их рассмотреть с точки зрения того или иного теоретического подхода.

На занятиях материал излагается в проблемной форме. Основной упор в преподавании делается на изучение теоретических понятий и возможности их применения на конкретных примерах, в том числе в устных выступлениях обучающихся.

Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся.

- Основные типы библиотек в R и их структура.
- Выполнение в R простейших операций, которые слушатели уже умеют на языке Python (ifelse, gsub, grepl и проч.)
- Практика объединений и преобразования датафреймов.
- Фильтрация и преобразование переменных с помощью функций ifelse, gsub и проч..
- Основы визуализации данных в ggplot2
- Преимущества data.table.
- Работа с разными типами apply family.
- Основные методы кластеризации и их применение в R: K-Means, Mean-Shift, DBSCAN.

Источники для самостоятельной подготовки:

- Введение в статистическое обучение с примерами на языке R: Практическое пособие / Д.Гарет и др.-М.: ДМК Пресс, 2017.-456 с.(П)
- Сбор данных в Интернете на языке R: Практическое руководство / Д.А.Храмов-М.: ДМК Пресс, 2017.-280 с.
- Богданов, Е. П. Интеллектуальный анализ данных : практикум для магистрантов направления 09.04.03 «Прикладная информатика» профиль подготовки «Информационные системы и технологии корпоративного управления» / Е. П. Богданов. - Волгоград : ФГБОУ ВО Волгоградский ГАУ, 2019. - 112 с. - Текст :

- электронный. - URL: <https://znanium.com/catalog/product/1087885> . – Режим доступа: по подписке.
- Барский, А. Б. Планирование виртуальных вычислений : учеб. пособие / А.Б. Барский. — М. : ИД «ФОРУМ» : ИНФРА-М, 2018. — 200 с. — (Высшее образование). — www.dx.doi.org/10.12737/19901. - ISBN 978-5-8199-0655-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/966062> . – Режим доступа: по подписке.
 - Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход/ЛемешкоБ.Ю., ЛемешкоС.Б., ПостоваловС.Н. и др. - Новосибирск : НГТУ, 2011. - 888 с.: ISBN 978-5-7782-1590-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/548140> . – Режим доступа: по подписке.

5. ОЦЕНОЧНЫЕ СРЕДСТВА

Проведение текущего контроля в рамках реализации данной дисциплины проходит в соответствии с Таблицей 3 данной рабочей программы дисциплины по основным понятиям (категориям) и проблемам, рассматриваемым в предложенных темах. Фиксация результатов текущего контроля в рамках реализации данной дисциплины не предусмотрена.

Типовые задания к текущей аттестации.

Тема 1. Введение.

Практическое задание 1:

- Установить R-Studio
- Установить и загрузить ряд библиотек
- Охарактеризовать типы данных в предложенном датасете
- Произвести простейшие преобразования с данными.

Тема 2. Основы R.

Практическое задание 2: •

- Создать новый датасет
- Объединить два датасета
- Изменить тип датафрейма с long2wide на wide2long.

Тема 3. Библиотека Dplyr.

Практическое задание 3:

- Выбрать нужные строки по заданному условию
- Создать новую колонку по заданному условию
- Рассчитать среднее для различных категорий в данных
- Пример: в данных есть информация по 3-м университетам. Вам предстоит рассчитать средние оценки и количество студентов для каждого ВУЗа.
- Визуализировать полученные результаты.

Тема 4 Data.table

Практическое задание 4: Произвести те же действия, что и в Разделе 3, но при помощи другой библиотеки в R.

Тема 5. Циклы, специфика apply family.

Практическое задание 5.

- Написать собственный цикл, который определяет четные и нечетные числа
- Применить функцию из семейства apply для преобразования данных
- Применить функцию lapply для преобразования данных типа list.

Тема 6. Методы кластеризации и кластеризация в R.

Практическое задание 6.

Выполнить основные виды кластеризации на учебном датасете.

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Практическое задание	к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено слушатель выполняет задание частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, полное и правильное выполнение задания в соответствии с требованиями

Форма промежуточной аттестации – зачет, выставляемый на основе устного опроса.

При аттестации используются система «зачтено» и «не зачтено» в соответствии с критериями оценивания.

В результате промежуточного контроля знаний обучающиеся получают аттестацию по дисциплине.

Показатели, критерии и оценивание компетенций по уровням их формирования в процессе промежуточной аттестации

Таблица 4

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Коды ЗУВ (в соответствии с Таблицей 1)	Критерии оценивания	Оценка
зачет / проект	ПК-5	З (ПК-5) У (ПК-5) В (ПК-5)	Слушатель демонстрирует полную самостоятельность в подборе фактического материала и аналитическое отношение к нему, умение рассматривать примеры и факты во взаимосвязи и взаимообусловленности, отбирать наиболее существенные из них; а также показывает грамотное использование методов описания и презентации исследования	зачтено
			Слушатель не демонстрирует аналитическое отношение к материалу, не видит взаимосвязь примеров и фактов; а также использует методы описания и презентации исследования с большим количеством существенных ошибок	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «зачтено», показывают уровень сформированности у обучающегося компетенций.

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «не зачтено», показывают не сформированность у обучающегося компетенций по дисциплине.

Типовые задания к промежуточной аттестации.

В качестве финального проекта слушателем будут предложены различные наборы данных над которыми необходимо будет совершить преобразования в соответствии с разобранными темами.

Пример задания (1): предоставить описательные статистики по датасету и представить в виде таблицы.

Пример задания (2): Визуализировать связь между двумя показателями и дать краткое описание полученной иллюстрации.

Критерии оценки:

1. Количество правильно выполненных задач.
2. Использование различных вариантов для выполнения задания (base, dplyr, data.table).
3. Корректная интерпретация написанного кода (Слушатель может объяснить, что выполняют функции в коде).

Практические задания выполняются в R-studio: импорт данных, манипуляцию с переменными, выбор оптимального метода для решения задачи, оформления результатов в табличной и графической форме, а также их интерпретацию.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

- Введение в статистическое обучение с примерами на языке R / Е. Джеймс, Д. Уитгон, Т. Хасти, Р.В. Тибширани ; пер. с англ. С.Э. Мاستицкого. - 2-е изд., испр. - Москва : ДМК Пресс, 2017. - 456 с. - ISBN 978-5-97060-495-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027867>
- Робинсон, Я. Графовые базы данных: новые возможности для работы со связанными данными / Ян Робинсон, Джим Вебер, Эмиль Эйфрем ; пер. с англ. Р.Н. Рагимова ; под науч. ред. А.Н. Кисилева. - 2-е изд. - Москва : ДМК Пресс, 2016. - 256 с. - ISBN 978-5-97060-201-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028068>
- Дуглас А., Л. Люк, Д.А. Анализ сетей (графов) в среде R. Руководство пользователя / Д.А. Люк ; пер. с англ. А.В. Груздева. - Москва : ДМК Пресс, 2017. - 250 с. - ISBN 978-5-97060-428-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/102809>
- Храмов, Д.А. Сбор данных в Интернете на языке R / Д. А. Храмов. - Москва : ДМК Пресс, 2017. - 280 с. - ISBN 978-5-97060-459-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028129>

Дополнительная литература:

- Нидхем, М. Графовые алгоритмы : практическое руководство / М. Нидхем, Э. Ходлер ; пер. с англ. В. С. Яценкова. - Москва : ДМК Пресс, 2020. - 258 с. - ISBN 978-5-97060-799-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1094928>
- Социально-экономическая статистика: Учебное пособие [Электронный ресурс] / Я.С. Мелкумов. - М.: НИЦ Инфра-М, 2013. - 236 с. - URL: <http://znanium.com/bookread2.php?book=366756>
- Общая теория статистики: Учебное пособие / С.Н. Лысенко, И.А. Дмитриева. - Изд., испр. и доп. - М.: Вузовский учебник: НИЦ ИНФРА-М, 2014. - 219 с. - [Электронный ресурс]. URL: <http://znanium.com/bookread2.php?book=397795>
- Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python : [16+] / В.М. Волкова, М.А. Семенова, Е.С. Четвертакова, С.С. Вожов ; Новосибирский государственный технический университет. – Новосибирск : Новосибирский государственный технический университет, 2017. – 74 с. : ил., табл. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=576496>

Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения программы:

Информационно-справочные системы:

- Гарант.Ру. Информационно-правовой портал: <http://www.garant.ru>
- Информационная система «Единое окно доступа к образовательным ресурсам»: <http://window.edu.ru/>
- Открытое образование. Ассоциация «Национальная платформа открытого образования»: <http://npoed.ru>
- Официальная Россия. Сервер органов государственной власти Российской Федерации: <http://www.gov.ru>
- Официальный интернет-портал правовой информации. Государственная система правовой информации: <http://pravo.gov.ru>
- Правовой сайт КонсультантПлюс: <http://www.consultant.ru/sys>
- Российское образование. Федеральный портал: <http://www.edu.ru>

Тематические системы:

- Google. Книги: <https://books.google.com>
- Internet Archive: <https://archive.org>
- Koob.ru. Электронная библиотека «Куб»: <http://www.koob.ru/philosophy/>
- Библиотека Ихтика [ihtik.lib.ru]: <http://ihtik.lib.ru/>
- Докусфера — Российская национальная библиотека: <http://leb.nlr.ru>
- ЕНИП — Электронная библиотека «Научное наследие России»: <http://e-heritage.ru/>
- Интелрос. Интеллектуальная Россия: <http://www.intelros.ru/>
- Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
- Неприкосновенный запас: <http://magazines.russ.ru/nz/>
- Президентская библиотека: <http://www.prlib.ru>
- Российская государственная библиотека: <http://www.rsl.ru/>
- Российская национальная библиотека: <http://www.nlr.ru/poisk/>

7. ПРОГРАММНОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММЫ

В ходе реализации образовательного процесса используются многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа и семинарского типа обеспечивается демонстрационным оборудованием.

Для обучающихся с ограниченными возможностями здоровья и инвалидов (в случае необходимости) могут быть созданы специальные условия для получения образования.

Программное обеспечение

При осуществлении образовательного процесса в рамках Университета слушателям рекомендовано использовать следующее лицензионное программное обеспечение:

- OS Microsoft Windows (OVS OS Platform)
- MS Office (OVS Office Platform)
- Adobe Acrobat Professional 11.0 MLP AOO License RU
- Adobe CS5.5 Design Standart Win IE EDU CLP
- ABBYY FineReader 11 Corporate Edition
- ABBYY Lingvo x5
- Adobe Photoshop Extended CS6 13.0 MLP AOO License RU
- Adobe Acrobat Reader DC /Pro – бесплатно
- Google Chrome – бесплатно

- Opera – бесплатно
- Mozilla – бесплатно
- VLC – бесплатно
- R – бесплатно