

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 31.10.2023 11:15:37

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e3191f380051f

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

Факультет экономики

УТВЕРЖДАЮ:

Ректор  В.В. Волков

« 04 » апреля 2022 г.

Протокол Ученого Совета
№ 3 от 30 марта 2022 г.

Рабочая программа дисциплины
Введение в R

образовательная программа
направление подготовки
38.04.01 Экономика

направленность (профиль)
«Экономика и финансы»
программа подготовки – магистратура

язык обучения – русский
форма обучения – очная

квалификация (степень) выпускника
Магистр

Санкт-Петербург

Автор:

Раскина Ю.В., кандидат экономических наук, доцент факультета экономики АНООВО «ЕУСПб»

Кучаков Руслан Камилович, младший научный сотрудник, Институт проблем правоприменения при АНООВО «ЕУСПб»

Рецензент:

Соколов М.В., кандидат экономических наук, старший научный сотрудник факультета экономики АНООВО «ЕУСПб»

Рабочая программа дисциплины «**Введение в Р**», входящей в образовательную программу уровня магистратуры «Экономика и финансы», утверждена на заседании Совета факультета экономики.

Протокол заседания № 7 от 07 марта 2022 года.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ
«Введение в R»
Б1.В.06

Дисциплина **«Введение в R»** является обязательной дисциплиной части, формируемой участниками образовательных отношений, Блока 1. «Дисциплины (модули)» основной профессиональной образовательной программы высшего образования «Экономика и финансы» по направлению подготовки 38.04.01 Экономика.

Дисциплина реализуется на факультете экономики.

Дисциплина нацелена на формирование следующих компетенций, которыми должен обладать выпускник, освоивший программу магистратуры:

профессиональных компетенций (ПК):

— способен анализировать и использовать различные источники информации для проведения экономических расчетов, готовить аналитические материалы для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне (ПК-3);

— способен работать с различными источниками статистической информации, разрабатывать и совершенствовать методологии сбора и обработки статистических данных, проводить статистические расчеты с применением соответствующих математических методов и информационных технологий, а также последующую аналитическую работу с полученными данными (ПК-5).

Дисциплина **«Введение в R»** является введением в программную среду R. Дисциплина нацелена на приобретение базовых навыков программирования в среде R для решения экономических и финансовых задач. Основные разделы дисциплины включают: понимание логики языка, освоение основных механик навигации и отбора наблюдений/переменных, освоение методов преобразования данных.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, практические занятия, самостоятельная работа магистранта, промежуточный контроль.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточный контроль в форме зачета с оценкой (в конце 2 модуля).

Общая трудоемкость освоения дисциплины составляет 2 зачетных единицы, 72 часа.

Программой дисциплины предусмотрены: 14 часов лекций, 14 часов практических занятий, 44 часа самостоятельной работы магистранта.

Содержание

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	6
4. ОБЪЕМ ДИСЦИПЛИНЫ	6
5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ	7
5.1 Содержание дисциплины	7
5.2 Структура дисциплины	8
6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ	9
6.1 Общие положения.....	9
6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины.....	9
6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине.....	10
6.4 Перечень литературы для самостоятельной работы обучающегося:.....	12
6.5 Перечень учебно-методического обеспечения для самостоятельной работы.....	12
7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ	12
7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации	12
7.2 Контрольные задания для текущей аттестации	14
7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации	18
7.4 Типовые задания к промежуточной аттестации	20
7.5 Средства оценки индикаторов достижения компетенций	22
8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	23
8.1. Основная литература	23
8.2. Дополнительная литература	23
9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	24
9.1 Программное обеспечение	24
9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:	24
9.3 Лицензионные электронные ресурсы библиотеки Университета	25
9.4 Электронная информационно-образовательная среда Университета.....	26
10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	26
ПРИЛОЖЕНИЕ 1	27

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины «Введение в R» заключается в знакомстве студентов с языком программирования R и приобретении базовых навыков работы в программной среде R для решения экономических задач.

Задачи дисциплины:

1. дать знания об основных элементах среды программирования R,
2. дать общее представление о логике языка R,
3. научить манипуляции и преобразованию данных с использованием языка R применительно к экономическим задачам,
4. научить основам графического отображения данных с использованием R применительно к экономическим данным.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины обучающийся должен овладеть следующими компетенциями: профессиональными (ПК). Планируемые результаты формирования компетенций и индикаторы их достижения в результате освоения дисциплины представлены в Таблице 1.

Таблица 1

Планируемые результаты освоения дисциплины, соотнесенные с индикаторами достижения компетенций обучающихся

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ПК-3 Способен анализировать и использовать различные источники информации для проведения экономических расчетов, готовить аналитические материалы для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне	ИД.ПК-3.1. Формирует необходимую информационную базу для исследовательского процесса, оценивает надежность информации	Знать: логику и специфику аналитических мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне; структуру и особенности различных источников информации для проведения экономических расчетов З (ПК-3)
	ИД.ПК-3.2. Осуществлять систематическое наблюдение за состоянием экономических процессов	
	ИД.ПК-3.3. Обобщает и анализировать данные и на основе полученных результатов обосновывать принятие стратегических и тактических решений на различных уровнях управления	
	ИД.ПК-3.4. Оценивает эффективность мероприятий в области экономической политики и стратегических решений	Уметь: осуществлять систематическое наблюдение за состоянием экономических процессов, обобщать и анализировать данные и на основе полученных результатов обосновывать принятие стратегических и тактических решений на различных уровнях управления У (ПК-3)
	ИД.ПК-3.5. Применяет аналитический инструментарий проведения экономических расчетов	Владеть: навыками систематизации и обобщения накопленных знаний для саморазвития и самореализации, инструментарием оценки эффективности мероприятий в области экономической политики и стратегических решений, аналитическим инструментарием проведения экономических расчетов, способностью принимать организационно-управленческие решения В (ПК-3)
ПК-5 Способен работать с различными источниками статистической информации	ИД.ПК-5.1. Работает с различными источниками статистической информации ИД.ПК-5.2. Производит экспериментальные статистические	Знать: российские и международные методологические положения и стандарты, методические подходы к проведению статистических расчетов и анализу,

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
информации, разрабатывать и совершенствовать методологии сбора и обработки статистических данных, проводить статистические расчеты с применением соответствующих математических методов и информационных технологий, а также последующую аналитическую работу с полученными данными	расчеты с применением информационных технологий ИД.ПК-5.3. Анализирует данные на основе методов математической статистики ИД.ПК-5.4. Разрабатывает и совершенствует системы статистических показателей и методик их расчета	статистические пакеты прикладных программ, актуальные научные публикации по статистике, в том числе зарубежные, источники статистической информации З (ПК-5)
		Уметь: производить статистические расчеты с применением соответствующих математических методов и информационных технологий, а также последующую аналитическую работу с полученными данными, применять статистические пакеты прикладных программ У (ПК-5)
		Владеть: навыками применения традиционных и инновационных методов статистического анализа В (ПК-5)

В результате освоения дисциплины магистрант должен:

ЗНАТЬ: основы языка программирования R; базовые принципы работы с данными с помощью R;

УМЕТЬ: проводить типовые манипуляции с данными, осуществлять подготовку данных к анализу и реализовывать вычисления в R;

ВЛАДЕТЬ: навыками типовых манипуляций с данными; навыками обработки и сбора данных; основами вычислений в R.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Введение в R» является обязательной дисциплиной части, формируемой участниками образовательных отношений, Блока 1 «Дисциплины (модули)» образовательной программы «Экономика и финансы». Код дисциплины по Учебному плану Б1.В.06. Курс читается во втором модуле, форма промежуточной аттестации – зачет с оценкой.

Для освоения данной дисциплины достаточно подготовки в области отдельных разделов высшей математики в объеме специалитета или бакалавриата по экономическим или физико-математическим специальностям (направлениям).

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе изучения следующих дисциплин, а также прохождения учебной и производственной практик:

Б1.В.10 Анализ временных рядов;

Б1.В.13 Модели микроэконометрики

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 2 (две) зачетных единицы, 72 часа.

Таблица 2

Типы учебных занятий и самостоятельная работа		Объем дисциплины										
		Всего	Модуль									
			1	2	3	4	5	6	7	8	9	10
<i>Очная форма обучения</i>												
Контактная работа обучающихся с преподавателем в соответствии с УП:		28	-	28	-	-	-	-	-	-	-	-
лекционного типа (Лек)		14	-	14	-	-	-	-	-	-	-	-
практические занятия (Пр)		14	-	14	-	-	-	-	-	-	-	-
Самостоятельная работа обучающихся (СР)		44	-	44	-	-	-	-	-	-	-	-
Промежуточная аттестация	форма	Зачет с оценкой	-	Зачет с оценкой	-	-	-	-	-	-	-	-
	час.	-	-	-	-	-	-	-	-	-	-	-
Общая трудоемкость (час. / з.е.)		72/2	-	72/2	-	-	-	-	-	-	-	-

5. 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) по средствам индикаторов достижения компетенций в соответствии с Таблицей 3.

5.1 Содержание дисциплины

Таблица 3

Содержание дисциплины					
№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)
1	Введение в среду R	Цели и задачи дисциплины, обзор основных тем и вопросов. Рабочее пространство R/RStudio. Глобальное окружение. Справка. Stackoverflow. Установка пакетов. Типы и структуры данных. Проверка типа данных (общие функции и семейство функций is..) и преобразования данных (семейство функций as..). Присвоение и создание объектов данных.	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	3 (ПК-3) У (ПК-3) В (ПК-3) 3 (ПК-5) У (ПК-5) В (ПК-5)
2	Базовые вычисления и управляющие конструкции	Базовые математические операции и статистические функции. Сравнение (if/ifelse). Циклы (for, while). Функции. Стили написания кода в R: базовый синтаксис v. tidyverse v. data.table.	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	3 (ПК-3) У (ПК-3) В (ПК-3) 3 (ПК-5) У (ПК-5) В (ПК-5)
3	Загрузка и сохранение данных	Работа с данными разных форматов: .csv/.xls/.xlsx/.dta/.sav/.rdata. Загрузка/чтение и сохранение данных в различных форматах.	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	3 (ПК-3) У (ПК-3) В (ПК-3) 3 (ПК-5) У (ПК-5) В (ПК-5)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)
4	Преобразование данных	Манипуляции с данными: поиск нужных значений и отбор. Основные виды трансформации и преобразования данных. Длинный и широкий формат таблиц. Отбор переменных и наблюдений. Создание и объединение таблиц.	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	3 (ПК-3) У (ПК-3) В (ПК-3) 3 (ПК-5) У (ПК-5) В (ПК-5)
5	Основные статистические функции	Реализация мер тенденции: среднее арифметическое, медианы, мода. Меры разброса: среднеквадратичное отклонение, дисперсия, интерквартильный размах. Плоские таблицы и частотные распределения. Эмпирическая функция распределения. Коэффициент корреляции Пирсона, Кендалла, Спирмена. Сравнение значений и средних.	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	3 (ПК-3) У (ПК-3) В (ПК-3) 3 (ПК-5) У (ПК-5) В (ПК-5)
6	Основы работы с текстовыми данными	Базовые функции для работы со строчными данными. Знакомство с регулярными выражениями.	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	3 (ПК-3) У (ПК-3) В (ПК-3) 3 (ПК-5) У (ПК-5) В (ПК-5)
7	Основы сбора данных (data mining; parsing)	Основы сбора данных из различных источников. Знакомство с файлами формата: .csv/.xls/.xlsx/.xml/.json. Знакомство с POST/GET запросами и работой с API (Application Programming Interface). Предобработка, проверка на целостность данных и подготовка данных к анализу.	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	3 (ПК-3) У (ПК-3) В (ПК-3) 3 (ПК-5) У (ПК-5) В (ПК-5)

5.2 Структура дисциплины

Таблица 4

Структура дисциплины							
№ п/п	Наименование тем (разделов)	Объем дисциплины, час.					Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП			СР	
			Лек	Лаб	Пр		
<i>Очная форма обучения</i>							
Тема 1	Введение в среду R	10	2	-	2	6	ДЗ
Тема 2	Базовые вычисления и управляющие конструкции	10	2	-	2	6	ДЗ
Тема 3	Загрузка и сохранение данных	10	2	-	2	6	ДЗ
Тема 4	Преобразование данных	10	2	-	2	6	ДЗ

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.					Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП			СР	
			Лек	Лаб	Пр		
Очная форма обучения							
Тема 5	Основные статистические функции	10	2	-	2	6	ДЗ
Тема 6	Основы работы с текстовыми данными	11	2	-	2	7	ДЗ
Тема 7	Основы сбора данных (data mining: parsing)	11	2	-	2	7	ДЗ
Промежуточная аттестация		-	-	-	-	-	Зачет с оценкой
Всего:		72/2	14	-	14	44	-

*Примечание: формы текущего контроля успеваемости: домашнее задание (ДЗ).

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

6.1 Общие положения

Знания и навыки, полученные в результате лекций и занятий семинарского типа, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения текстов и исследовательской литературы (из списков основной и дополнительной литературы) и их анализа.

Самостоятельная работа является важнейшей частью процесса высшего образования. Ее следует осознанно организовать, выделив для этого необходимое время и соответствующим образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/практическим (семинарским) занятиям. Литературу, рекомендованную в программе курса, следует, по возможности, читать в течение всего модуля, концентрируясь на обусловленных программой курса темах.

Существенную часть самостоятельной работы магистранта представляет самостоятельное изучение вспомогательных учебно-методических изданий, лекционных конспектов, интернет-ресурсов и пр. Подготовка к практическим занятиям является важной формой работы магистранта. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя.

6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины

Тема 1. Введение в среду R.

1.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 3 часа.

1.2. Подготовка к практическим занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 3 часа.

Итого: 6 часов.

Тема 2. Базовые вычисления и управляющие конструкции.

2.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 3 часа.

2.2. Подготовка к практическим занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 3 часа.
Итого: 6 часов.

Тема 3. Загрузка и сохранение данных.

3.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 3 часа.

3.2. Подготовка к практическим занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 3 часа.
Итого: 6 часов.

Тема 4. Преобразование данных.

4.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 3 часа.

4.2. Подготовка к практическим занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 3 часа.
Итого: 6 часов.

Тема 5. Основные статистические функции.

5.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 3 часа.

5.2. Подготовка к практическим занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 3 часа.
Итого: 6 часов.

Тема 6. Основы работы с текстовыми данными.

6.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 3 часа.

6.2. Подготовка к практическим занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа.
Итого: 7 часов.

Тема 7. Основы сбора данных (data mining: parsing).

7.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 3 часа.

7.2. Подготовка к практическим занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа.
Итого: 7 часов.

6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине

1. Введение в среду R

1.1 Назовите основные типы данных, поддерживаемые R. Чем они характеризуются?

1.2 Назовите основные виды объектов данных в R. Чем они характеризуются?

1.3 Как узнать и изменить тип данных выбранного объекта из окружения?

1.4 Чем отличается матрица от data.frame?

2. Базовые вычисления и управляющие конструкции

2.1 Назовите основные отличия конструкций if-else/ifelse.

2.2 Чем отличается цикл «for» и «while»?

2.3 Предложите схему решения следующей задачи: необходимо перебрать вектор поэлементно и заменить все нулевые значения на NA.

2.4 Предложите схему решения следующей задачи: необходимо написать функцию, принимающую два вектора одинаковой размерности с произвольным набором элементов, произвести деление на 100. Функция должна возвращать полученный результат в виде нового вектора.

3. Загрузка и сохранение данных

3.1 Если вам нужно загрузить данные в виде плоской таблицы с неизвестным разделителем, как вы поступите?

3.2 Опишите схему решения следующей задачи: вам нужно проигнорировать при загрузке n первых строк и отобрать только нужные переменные.

3.3 Как вы думаете, в чем могут состоять риски использования при загрузке неизвестного датасета типов данных, определяемых по умолчанию?

3.4 Опишите схему решения следующей задачи: как сохранить данные с новым базовым разделителем в формате плоской таблицы?

4. Преобразование данных

4.1 Чем отличается «длинный» формат представления табличных данных от «широкого»?

4.2 Как преобразовать «длинный» формат в «широкий» и наоборот?

4.3 Опишите схему решения следующей задачи: у вас есть два исходных датасета. Первый содержит данные экономических показателей субъектов РФ в годовом разрезе в «широком» формате. Второй датасет содержит информацию о количестве юридических лиц, зарегистрированных в субъектах РФ в годовом разрезе. Данные представлены в длинном формате. Вам необходимо объединить два датасета по переменным регион-год.

5. Основные статистические функции

5.1 Назовите функции R с реализацией расчетов дисперсии, стандартного отклонения, арифметического среднего, медианы.

5.2 Как построить доверительные интервалы для среднеарифметического?

5.3 Назовите функцию для сравнения разницы средних?

6. Основы работы с текстовыми данными

6.1 Опишите способ детекции искомого паттерна в текстовых данных.

6.2 Опишите риски ошибок первого и второго рода при использовании поисковых запросов для выделения текстовых данных (regex).

6.3 Каким образом возможно осуществить замену определенного паттерна на указанные значения в текстовых данных?

7. Основы сбора данных (data mining: parsing)

7.1 Назовите разделитель колонок в стандартном файле формата .csv.

7.2 Опишите структуру модели XML файла.

7.3 Перечислите возможные способы проверки данных на целостность.

6.4 Перечень литературы для самостоятельной работы обучающегося:

1. Митчелл, Р. Скрапинг веб-сайтов с помощью Python / Р. Митчелл ; пер. с англ. А. В. Груздева. - Москва : ДМК Пресс, 2016. - 280 с. - ISBN 978-5-97060-223-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027754>
2. Солем, Ян Эрик Программирование компьютерного зрения на языке Python / Ян Эрик Солем ; пер. с англ. А.А. Слинкина. - Москва : ДМК Пресс, 2016. - 312 с. - ISBN 978-5-97060-200-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027847>
3. Маккинни, У. Python и анализ данных : практическое пособие / У. Маккинни ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2020. - 540 с. - ISBN 978-5-97060-590-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2012523>
4. Коэльо, Луис Педро Построение систем машинного обучения на языке Python / Луис Педро Коэльо, Вилли Ричарт ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2016. - 302 с. - ISBN 978-5-97060-330-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027824>
5. Рамальо, Л. Python. К вершинам мастерства / Лучано Рамальо ; пер. с англ. А.А. Слинкина. - Москва : ДМК Пресс, 2016. - 768 с. - ISBN 978-5-97060-384-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028052>
6. Вестра, Э. Разработка геоприложений на языке Python / Э. Вестра ; пер. с англ. А.В. Логунова. - Москва : ДМК Пресс, 2017. - 446 с. - ISBN 978-5-97060-437-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028127>
7. Мартин, О. Байесовский анализ на Python : практическое пособие / О. Мартин ; пер. с англ. А. В. Снастина. - Москва : ДМК Пресс, 2020. - 340 с. - ISBN 978-5-97060-768-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1094942>
8. 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>

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «Введение в R» разработано учебно-методическое обеспечение в составе:

1. Контрольные задания для подготовки к процедурам текущего контроля (п. 7.2 Рабочей программы).
2. Типовые задания для подготовки к промежуточной аттестации (п. 7.4 Рабочей программы).
3. Рекомендуемые основная, дополнительная литература, Интернет-ресурсы и справочные системы (п. 8, 9 Рабочей программы).
4. Рабочая программа дисциплины размещена в электронной информационно-образовательной среде Университета на электронном учебно-методическом ресурсе АНООВО «ЕУСПб» — образовательном портале LMS Sakai — Sakai@EU.

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся

научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому занятию, активное слушание на лекциях, выполнение домашних заданий. Магистрант должен присутствовать на лекциях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по темам обсуждения.

Текущий контроль проводится в форме оценивания выполнения домашних заданий, демонстрирующих степень знакомства магистрантов с дополнительной литературой.

Таблица 5

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Введение в среду R	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	З (ПК-3) У (ПК-3) В (ПК-3) З (ПК-5) У (ПК-5) В (ПК-5)	Домашние задания	зачтено/ не зачтено
Базовые вычисления и управляющие конструкции	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	З (ПК-3) У (ПК-3) В (ПК-3) З (ПК-5) У (ПК-5) В (ПК-5)	Домашние задания	зачтено/ не зачтено
Загрузка и сохранение данных	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	З (ПК-3) У (ПК-3) В (ПК-3) З (ПК-5) У (ПК-5) В (ПК-5)	Домашние задания	зачтено/ не зачтено
Преобразование данных	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	З (ПК-3) У (ПК-3) В (ПК-3) З (ПК-5) У (ПК-5) В (ПК-5)	Домашние задания	зачтено/ не зачтено
Основные статистические функции	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	З (ПК-3) У (ПК-3) В (ПК-3) З (ПК-5) У (ПК-5) В (ПК-5)	Домашние задания	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Основы работы с текстовыми данными	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	З (ПК-3) У (ПК-3) В (ПК-3) З (ПК-5) У (ПК-5) В (ПК-5)	Домашние задания	зачтено/ не зачтено
Основы сбора данных (data mining: parsing)	ПК-3 ПК-5	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	З (ПК-3) У (ПК-3) В (ПК-3) З (ПК-5) У (ПК-5) В (ПК-5)	Домашние задания	зачтено/ не зачтено

При освоении дисциплины домашние задания оцениваются с использованием 100-балльной шкалы с последующим переводом в бинарную систему для получения результатов текущего контроля, фиксирующих ход образовательного процесса, фиксирующих ход образовательного процесса, согласно Таблице 6.

Таблица 6

Критерии оценивания

Формы текущего контроля успеваемости	Описание	Показатели	Количество баллов по 100-балльной шкале	Результаты текущего контроля
Домашнее задание	Выполняя домашнее задание, состоящее в письменном изложении ответов на теоретические вопросы и/или решений практических задач, предложенных преподавателем, обучающиеся закрепляют пройденный материал и готовятся к усвоению нового.	Обучающийся правильно выполняет задания и отвечает на все поставленные в задании вопросы.	41–100	зачтено
		Обучающийся испытывает затруднения при выполнении предложенных заданий.	0–40	не зачтено

7.2 Контрольные задания для текущей аттестации

Примеры вопросов домашнего задания

Тема 1. Введение в среду R

1.1 Сохраните названия колонок в отдельный вектор
vector_names <- ...

1.2 Удвойте количество имен в созданном векторе vector_names, сохраните полученный результат в vector_names_doubled
vector_names_doubled <- ...

1.3 Удалите из вектора vector_names имена "Ozone" и "Month", сохраните полученный результат в vector_names_resaved

```

vector_names_resaved <- ...
# 1.4 Вызовите переменную "Solar.R" тремя разными способами
first <- ...
second <- ...
third <- ...
# 1.5 Вызовите первые 10 строк датасета двумя разными способами
airquality_head_ten_rows_1 <- ...
airquality_head_ten_rows_2 <- ...
# 1.6 Вызовите последние 10 строк датасета двумя разными способами
airquality_tail_ten_rows_1 <- ...
airquality_tail_ten_rows_2 <- ...
# 1.7 Создайте новый датасет из переменных "Wind" и "Temp"
airquality_selected <- ...
# 1.8 Удалите из окружения объекты
first/second/third/vector_names/vector_names_doubled/vector_names_resaved

## Операторы
# 1.9 Выведете все строки, где переменная "Ozone" больше 100
airquality_ozon_more_than_100 <- airquality[airquality$Ozone...]
# 1.10 Избавимся от пропусков в данных, используя ф-ю !is.na().
# Уберите все пропущенные значения и пересохраните DF
airquality_ozon_more_than_100 <- airquality_ozon_more_than_100[...]
# 1.11 Объединим две операции в одной строке с помощью оператора "И" — "&"
airquality_ozon_more_than_100 <- ...
# 1.12 Отберем наблюдения за сентябрь со значением переменной "Ozone" 50 и более
airquality_sep_ozon_at_least_20 <- airquality[airquality$Month == 9 & airquality$Ozone >
50,]
# 1.13 Отберете все непустые строки и все наблюдения за все месяцы кроме сентября
airquality_remove_sep <- ...

```

Тема 2. Базовые вычисления и управляющие конструкции

```

## 2.1 Напишите цикл «for» с перебором имен переменных.
## 2.2 Напишите цикл «for» с последовательным попарным суммированием элементов
двух векторов равной размерности с произвольными значениями. Результат: вектор,
содержащий попарную сумму элементов исходных векторов.
## 2.3 Используя конструкцию if-else/ifelse, создайте правило отбора всех элементов
вектора «больше»/«меньше»/«больше или равно»/«меньше или равно».
## 2.4 Используя конструкцию if-else/ifelse, создайте правило отбора всех строчных
элементов вектора, соответствующих выборному значению.

```

Тема 3. Загрузка и сохранение данных

```

# используя функцию read.csv:
# 3.1 загрузите только первые три колонки из предложенного датасета
# 3.2 загрузите первые 100 строк
# 3.3 используя аргумент na.strings, замените отсутствующие значения, на NA
# 3.4 проигнорируйте первые 20 строк при загрузке данных

```

Тема 4. Преобразование данных

```

## 4. Навигация и преобразование
# 4.1 Отберите все наблюдения, где переменная "Murder" принимает значения больше 10,
# а "UrbanPop" меньше 60. Сохраните результат в отдельный объект
usarrests_murder10_pop60 <- USArrests[...]

```

```

# 4.2 Используя DT с предыдущего шага, отберите и сохраните только первую и третью
строки
usarrests_murder10_pop60_first_and_third_rows<- usarrests_murder10_pop60[...]

# 4.3 Из исходного DT отберите все наблюдения, где переменная "Assault" больше или равна
200
# или переменная "Rape" принимает значения больше 20
usarrests_assault200_rape20<- USArrests[...]

# 4.4 Отберите наблюдения штатов "Ohio", "Texas" и "Vermont"
usarrests_states<- USArrests[...]

# 4.5 Найдите и сохраните штат-лидер по количеству убийств
usarrests_leader_murder_state<- USArrests[...]

# 4.6 Найдите и сохраните самый безопасный штат, с точки зрения кол-во убийств
usarrests_safest_murder_state<- USArrests[...]

# 4.7 Найдите и сохраните топ-5 самых опасных штатов по количеству нападений
# (при необходимости можно реализовать в несколько действий)
usarrests_top5_dangerous_states<- USArrests[...]

# 4.8 Найдите и сохраните топ-5 самых безопасных штатов по количеству нападений
# (при необходимости можно реализовать в несколько действий)
usarrests_top5_safest_states<- USArrests[...]

# 4.9 Используя датасет "usarrests_assault200_rape20", удалите колонки "Murder" и
"UrbanPop"
usarrests_assault200_rape20[...]

# 4.10 Данные USArrests, как следует из справки ?USArrests, содержат данные о количестве
# преступлений в расчёте на 100 тыс. населения. Пересчитайте показатель убийств "Murder"
# в расчёте на 1 тыс. и сохраните полученные данные в том же датасете в переменной
# murder_per_1000
USArrests[...]

# 4.11 Используя функции set.. упорядочите по возрастанию переменной "Murder" исходный
датасет
USArrests

# 4.12 Измените порядок следования колонок на: rn — UrbanPop — Assault — Rape —
Murder
USArrests

# 4.13 Переименуйте переменную "rn" в "states"
USArrests

```

Тема 5. Основные статистические функции

```

# 0. Вызовите первые 7 строк
swiss
# путь data.table

```



```

# 1. Отобразите связь между переменными Catholic и Fertility
ggplot(...) + geom_point()

# 2. Отобразите связь между переменными Education и Fertility
ggplot(...)

# 3. Отобразите связь между переменными Agriculture и Education
...

# 4.1 Посчитайте коэффициент корреляции Пирсона для трех пар переменных: Catholic и
Fertility, Education и Fertility, Agriculture и Education
# Catholic и Fertility
cor(...)

# 4.2. Education и Fertility
...

# 4.3. Agriculture и Education
...

# 5. Используя ф-ю cor и corrplot из пакета corrplot, отобразите попарные корреляции между
всеми переменными
cor_matrix <- cor(swiss)
...

# 6. Рассчитайте средние значения переменной Fertility при группировке по переменной
Catholic_half
swiss[, ..., by = ...]

# 7.1. Используя бинарную переменную Catholic_half, которая принимает значение равное
1,
# если доля католиков среди населения составляет более 50%, и 0 — во всех иных случаях,
# в качестве переменной группировки сравните средние переменных Fertility, Agriculture,
Education
t.test(data = , ...)

# 7.2. Fertility. Установите доверительный интервал на уровне 90%
t.test(...)

# 7.3. Education. Установите доверительный интервал на уровне 99% и выполните тест при
условии равенства дисперсий
t.test(...)

# 8. Выведите совместное распределение переменных fertility_high_than_median и
education_high_than_median
table(...)
# альтернатива
swiss[, .N, by = .(var1, var2)]

```

Тема 6. Основы работы с текстовыми данными

Создайте новые переменные: ИНН, ОГРН, ОКОГУ (код), ОКТМО, ОКОПФ (код), ОКАТО., используя в качестве основы предложенный текстовый документ: (1) извлеките нужные значения, (2) создайте из наборов значений соответствующие вектора, (3) объедините вектора в плоскую таблицу.

```
data.table(
```

```
inn = regmatches(text1, gregexpr("...")),
```

```
ogrn = ...
```

```
...
```

```
)
```

Тема 7. Основы сбора данных (data mining: parsing)

Создание панели на данных API

1. Используя готовые решения API для баз данных Всемирного банка и ВОЗ, загрузите данные за произвольный период времени и набор стран.
2. Каждый из вас должен загрузить два индикатора: первый, указан напротив вашей фамилии, второй – коллеги снизу. Последний в перечне в качестве второго индикатора обращается к значению первой ячейки.
3. Объедините два индикатора по стране и году, тем самым создав панель: страна–год–индикатор (1 и 2).
4. Оставьте в вашей панели данные только за последние 10 лет.

7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – **зачет с оценкой**, при выставлении которого учитываются результаты текущего контроля успеваемости обучающегося по дисциплине и результаты оценивания письменной зачетной работы, выраженные в 100-балльной шкале.

Зачетная работа – важнейший вид самостоятельной работы студентов, представляющий собой письменную работу по содержанию учебной дисциплины. Предполагается, что ход решения задач сопровождается подробными комментариями обучающегося.

Перед письменной зачетной работой проводится консультация, на которой преподаватель отвечает на вопросы обучающихся.

Критерии оценивания письменной зачетной работы представлены в таблице 7.

Таблица 7

Критерии оценивания письменной зачетной работы

Вид промежуточной аттестации	Показатели	Количество баллов
Письменная зачетная работа	Обучающийся глубоко и прочно усвоил программный материал; исчерпывающе, последовательно, четко и логически стройно излагает его в письменной зачетной работе, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами знаний.	81–100
	Обучающийся твердо знает программный материал, грамотно и по существу излагает его в письменной зачетной работе, не допуская существенных неточностей при выполнении заданий, правильно применяет теоретические положения при решении	61–80

Вид промежуточной аттестации	Показатели	Количество баллов
	практических задач, владеет необходимыми навыками и приемами их выполнения.	
	Обучающийся имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности при выполнении заданий письменной зачетной работы.	41–60
	Обучающийся не знает значительной части программного материала, допускает существенные ошибки, испытывает затруднения при выполнении заданий письменной зачетной работы..	0–40

Максимальная оценка по итогам освоения дисциплины составляет 100 баллов (80% оценки составляют результаты за домашние задания, 20% оценки составляют результаты письменной зачетной работы). При вычислении оценки обучающегося по итогам освоения дисциплины используется процедура округления до целого.

В результате промежуточного контроля знаний студенты получают аттестацию по дисциплине. На основании оценки обучающегося по итогам освоения дисциплины, выраженной в 100-балльной шкале, выставляется **зачет с оценкой** в соответствии с Таблицей 8.

Таблица 8

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Оценка по итогам освоения дисциплины (в 100-балльной шкале)	Результаты текущего контроля
зачет с оценкой / письменная зачетная работа	ПК-3 ПК-5	ИД.ПК-3.1.	З (ПК-3)	81–100	Зачтено, отлично
		ИД.ПК-3.2.	У (ПК-3)	61–80	Зачтено, хорошо
		ИД.ПК-3.3. ИД.ПК-3.4.	В (ПК-3) З (ПК-5)	41–60	Зачтено, удовлетворительно
		ИД.ПК-3.5. ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	У (ПК-5) В (ПК-5)	0–40	Не зачтено, неудовлетворительно

Результаты промежуточной аттестации по дисциплине во втором модуле, выраженные в оценках «зачтено, удовлетворительно», «зачтено, хорошо», «зачтено, отлично», свидетельствуют о сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Экономика и финансы» по направлению подготовки 38.04.01 Экономика (уровень магистратуры).

Результат промежуточной аттестации по дисциплине во втором модуле, выраженный в оценке «не зачтено, неудовлетворительно», свидетельствует об отсутствии или критическом уровне сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Экономика и финансы» по направлению подготовки 38.04.01 Экономика (уровень магистратуры).

7.4 Типовые задания к промежуточной аттестации

Список вопросов для подготовки к зачету с оценкой

Блок 1

1. Назовите основные типы данных, поддерживаемых R?
2. Назовите основные объекты данных, поддерживаемых R?
3. Как осуществить линейные преобразования над векторами в R?
4. Назовите основные отличия матрицы от data.frame

Блок 2

1. Назовите основные отличия цикла for от while.
2. Как использовать индексацию наблюдений для селекции строк. Назовите два различных способа.
3. Как вызвать все четные/нечетные элементы вектора?
4. Можно ли в data.frame хранить данные разных типов?

Блок 3

1. В чем заключаются риски использования назначаемых по умолчанию типов данных для переменных при загрузке неизвестного датасета?
2. Как заменить пропуски данных при загрузке на выбранные значения?
3. Как сохранить данные в формате csv с измененным основным разделителем?
4. Как сохранить данные в формате xlsx?

Блок 4

1. Чем отличается «широкий» формат представления табличных данных от «длинного»? Приведите примеры обоих видов отображения данных.
2. Назовите отличия data.frame от data.table.
3. Как обратиться к переменной заключенной в list и размещенной в data.frame?
4. Как объединить несколько data.frame?

Примеры заданий для письменной зачетной работы.

Вариант 1

1. Скачайте данные по предложенной ссылке.

1. Базовые навыки

```
# 1.1 Вызовите имена переменных датасета number_firms_spb  
number_firms_spb
```

```
# 1.2 Вызовите с 100 по 110 строки в датасете number_firms_spb  
number_firms_spb
```

```
# 1.3 Вызовите последние 10 строк датасета number_firms_spb  
number_firms_spb
```

```
# 1.4 Определите тип переменных датасета okved_codes_names  
okved_codes_names
```

```
# 1.5 Создайте новую переменную okved_factor в датасете okved_codes_names, которая  
будет идентична переменной okved_osn_code_full, но при этом являться факторной  
okved_codes_names
```

```
# 1.6 Отбери́ти из датасета number_firms_spb все наблюдения (коды ОКВЭД - год) с частотой более 2000 и сохраните их в отдельном датасете
number_firms_spb2k <- number_firms_spb
```

```
# 1.7 Найдите самый частотный код ОКВЭД во всем датасете number_firms_spb2k
number_firms_spb2k
```

```
# 1.8 Вызовите все наблюдения (ОКВЭД - год) из датасета number_firms_spb2k с частотой больше средней по всему датасету
number_firms_spb2k
```

```
# 1.9 Вызовите все наблюдения (ОКВЭД - год) из датасета number_firms_spb2k с частотой #превышающей медианное значение
number_firms_spb2k
```

2. Трансформация

```
# 2.1 Добавьте колонку с названиями кодов ОКВЭД (okved_osn_name_full) из датасета okved_codes_names в данные number_firms_spb2k
number_firms_spb2k
```

```
# 2.2 Просуммируйте данные number_firms_spb2k по годам и сохраните результат в number_firms_year
number_firms_year <- ...
number_firms_year
```

```
# 2.3 Используя ggplot отрисуйте тренд данных number_firms_year, где на оси X отложен год, а на оси Y – кол-во организаций
number_firms_year
```

```
# 2.4 Преобразуйте датасет number_firms_spb2k в широкий формат (ф-я dcast), # где по строкам будут значения кодов ОКВЭД, в колонках — года с 2010 по 2020, в пересечении — кол-во организаций.
# Сохраните полученный результат в number_firms_wide
```

```
number_firms_wide <- dcast(...)
number_firms_wide
```

```
# 2.5 Преобразуйте number_firms_wide из пункта 2.4 обратно в длинный формат (ф-я melt)
number_firms_molten
```

```
# 2.6 Используя регулярные выражения, отберите все наблюдения, содержащие морфемы, относящиеся к сфере торговли.
# Здесь и далее по раздлеу используем датасет number_firms_molten
...
```

```
# 2.7 Используя регулярные выражения, отберите все наблюдения, НЕ содержащие морфемы, относящиеся к сфере торговли
...
```

```
# 2.8 Используя регулярные выражения, удалите точки, разделяющие разделы ОКВЭД
...
```

2.9 Используя регулярные выражения, выделите («отрежьте») первые два знака (класс) в каждом коде ОКВЭД

...

2.10 Опираясь на результат из пункта 2.9, используйте полученную переменную для группировки, и вычислите среднее значение

...

3. Сравнение

3.1 Отобразите связь между значениями переменных "2011" и "2019" из данных number_firms_wide, используя ggplot

Используйте `2011`, `2019` для ввода числовых значений, когда они обозначают названия переменных

number_firms_wide

3.2 Посчитайте коэффициент корреляции Пирсона, используя данные number_firms_wide, между значениями переменных "2014" и "2019"

Используйте `2011`, `2019` для ввода числовых значений, когда они обозначают названия переменных

number_firms_wide

3.3 Посчитайте коэффициент корреляции Кендолла, используя данные number_firms_wide, между значениями переменных "2014" и "2019"

Используйте `2011`, `2019` для ввода числовых значений, когда они обозначают названия переменных

number_firms_wide

Вариант 2

2.1 Скачайте данные по предложенной ссылке.

2.2 Проверьте тип переменных, содержащихся в данных,

2.3 Присвойте наиболее подходящие типы данных взамен предложенных по умолчанию.

2.4 Преобразуйте данные в «длинный» формат.

2.5 Отберите наблюдения, содержащие в указанной переменной определенные значения.

2.6 Сохраните как отдельный .csv файл полученный результат.

7.5 Средства оценки индикаторов достижения компетенций

Таблица 9

Средства оценки индикаторов достижения компетенций

Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Средства оценки (в соот. с Таблицами 5, 7)
ПК-3	ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5.	Домашнее задание (ДЗ), письменная зачетная работа (ПЗР)
ПК-5	ИД.ПК-5.1. ИД.ПК-5.2. ИД.ПК-5.3. ИД.ПК-5.4.	Домашнее задание (ДЗ), письменная зачетная работа (ПЗР)

Таблица 10

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
Домашнее задание	<p>Магистрант в ходе подготовки и выполнения домашних заданий по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получивший развитие в рамках данной дисциплины:</p> <ol style="list-style-type: none"> 1. Анализирует различные источники информации для проведения экономических расчетов, готовит аналитические материалы для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне 2. Применяет математические методы и информационные технологии, проводит статистические расчеты, работает с различными источниками статистической информации, ведет аналитическую работу с данными, разрабатывает и совершенствует методологии сбора и обработки статистических данных
Письменная зачетная работа	<p>Магистрант в ходе подготовки и выполнения письменной зачетной работы, показывает способность совершать следующий набор профессиональных действий, получивший развитие в рамках данной дисциплины:</p> <ol style="list-style-type: none"> 1. Анализирует различные источники информации для проведения экономических расчетов, готовит аналитические материалы для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне 2. Применяет математические методы и информационные технологии, проводит статистические расчеты, работает с различными источниками статистической информации, ведет аналитическую работу с данными, разрабатывает и совершенствует методологии сбора и обработки статистических данных

8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

8.1. Основная литература

1. 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>

8.2. Дополнительная литература

1. Митчелл, Р. Скрапинг веб-сайтов с помощью Python / Р. Митчелл ; пер. с англ. А. В. Груздева. - Москва : ДМК Пресс, 2016. - 280 с. - ISBN 978-5-97060-223-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027754>

2. Солем, Ян Эрик Программирование компьютерного зрения на языке Python / Ян Эрик Солем ; пер. с англ. А.А. Слинкина. - Москва : ДМК Пресс, 2016. - 312 с. - ISBN 978-5-97060-200-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027847>

3. Маккинни, У. Python и анализ данных : практическое пособие / У. Маккинни ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2020. - 540 с. - ISBN 978-5-97060-590-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2012523>

4. Коэльо, Луис Педро Построение систем машинного обучения на языке Python / Луис Педро Коэльо, Вилли Ричарт ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2016. - 302 с. - ISBN 978-5-97060-330-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027824>

5. Рамальо, Л. Python. К вершинам мастерства / Лучано Рамальо ; пер. с англ. А.А. Слинкина. - Москва : ДМК Пресс, 2016. - 768 с. - ISBN 978-5-97060-384-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028052>

6. Вестра, Э. Разработка геоприложений на языке Python / Э. Вестра ; пер. с англ. А.В. Логунова. - Москва : ДМК Пресс, 2017. - 446 с. - ISBN 978-5-97060-437-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028127>

7. Мартин, О. Байесовский анализ на Python : практическое пособие / О. Мартин ; пер. с англ. А. В. Снастина. - Москва : ДМК Пресс, 2020. - 340 с. - ISBN 978-5-97060-768-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1094942>

9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

9.1 Программное обеспечение

При осуществлении образовательного процесса магистрантами и профессорско-преподавательским составом используется следующее лицензионное программное обеспечение:

1. OS Microsoft Windows (OVS OS Platform)
2. MS Office (OVS Office Platform)
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. ABBYY FineReader 11 Corporate Edition
6. ABBYY Lingvo x5
7. Adobe Photoshop Extended CS6 13.0 MLP AOO License RU
8. Adobe Acrobat Reader DC /Pro – бесплатно
9. Google Chrome – бесплатно
10. Opera – бесплатно
11. Mozilla – бесплатно
12. VLC – бесплатно
13. R — бесплатно
14. Python — бесплатно

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

Информационно-справочные системы

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

Профессиональные базы данных информационно-телекоммуникационной сети «Интернет»:

1. Google. Книги: <https://books.google.com>
2. Internet Archive: <https://archive.org>
3. Koob.ru. Электронная библиотека «Куб»: <http://www.koob.ru/philosophy/>
4. Библиотека Гумер – гуманитарные науки: <http://www.gumer.info>
5. Библиотека Ихтика [ihtik.lib.ru]: <http://ihtik.lib.ru/>
6. Докусфера — Российская национальная библиотека: <http://leb.nlr.ru>
7. ЕНИП — Электронная библиотека «Научное наследие России»: <http://e-heritage.ru/>

8. Интелрос. Интеллектуальная Россия: <http://www.intelros.ru/>
9. Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
10. Президентская библиотека: <http://www.prlib.ru>
11. Российская государственная библиотека: <http://www.rsl.ru/>
12. Российская национальная библиотека: <http://www.nlr.ru/poisk/>

9.3 Лицензионные электронные ресурсы библиотеки Университета

Профессиональные базы данных:

1. **Cambridge University Press** — полнотекстовая коллекция журналов издательства Cambridge University Press: <https://www.cambridge.org;>
2. **East View** – 100 ведущих российских журналов по гуманитарным наукам (архив и текущая подписка): <https://dlib.eastview.com/browse;>
3. **Ebook Central** коллекция электронных книг **Academic Complete** библиотеки компании **ProQuest** — **Ebook Central** — более 140 тыс. электронных научных книг крупнейших издательств мира: <https://ebookcentral.proquest.com;>
4. **EBSCO** – научные журналы, справочники, полнотекстовые и многопрофильные базы данных: <http://search.ebscohost.com;>
5. **eLIBRARY.RU** — Российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, наукометрическая база данных: <http://elibrary.ru;>
6. **JSTOR** – полнотекстовая база данных междисциплинарного характера, включающая более тысячи научных журналов по гуманитарным, социальным наукам и математике с их первого выпуска: [http://www.jstor.org/;](http://www.jstor.org/)
7. **Oxford Reference Online** — словари издательства Oxford University Press - [http://www.oxfordreference.com/;](http://www.oxfordreference.com/)
8. **Oxford University Press** — полнотекстовая коллекция журналов издательства Oxford University Press (текущая подписка и архив): [http://www.oxfordjournals.org/en/;](http://www.oxfordjournals.org/en/)
9. **Project MUSE Standard Collection** — полные тексты более чем 300 журналов по гуманитарным наукам зарубежных научных издательств: [http://muse.jhu.edu/;](http://muse.jhu.edu/)
10. **ProQuest Dissertations & Theses** — база диссертаций и дипломных работ: <http://search.proquest.com/pqdtglobal/dissertations;>
11. **Sage** — полнотекстовая коллекция журналов издательства Sage (текущая подписка и архив): [http://online.sagepub.com/;](http://online.sagepub.com/)
12. **SCOPUS** – реферативная наукометрическая база данных: <https://www.scopus.com;>
13. **Taylor&Francis** – полнотекстовая коллекция журналов издательства Taylor&Francis (текущая подписка и архив) – [http://www.tandfonline.com/;](http://www.tandfonline.com/)
14. **Web of Science** — реферативная наукометрическая база данных: <http://apps.webofknowledge.com;>
15. **Университетская информационная система РОССИЯ** — база электронных ресурсов для учебных программ и исследовательских проектов в области социально-гуманитарных наук: [http://www.uisrussia.msu.ru/;](http://www.uisrussia.msu.ru/)
16. Электронные журналы по подписке (текущие номера научных зарубежных журналов)

Электронные библиотечные системы:

1. **Znanium.com** – Электронная библиотечная система (ЭБС) – [http://znanium.com/;](http://znanium.com/)
2. Университетская библиотека онлайн – Электронная библиотечная система (ЭБС) – <http://biblioclub.ru/>

9.4 Электронная информационно-образовательная среда Университета

Образовательный процесс по дисциплине поддерживается средствами электронной информационно-образовательной среды Университета, которая включает в себя электронный учебно-методический ресурс АНООВО «ЕУСПб» — образовательный портал LMS Sakai — Sakai@EU, лицензионные электронные ресурсы библиотеки Университета, официальный сайт Университета (Европейский университет в Санкт-Петербурге [<https://euspr.org>]), локальную сеть и корпоративную электронную почту Университета, и обеспечивает:

- доступ к учебным планам, рабочим программам дисциплин (модулей), практик и к изданиям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочих программах;

- фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;

- формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок за эти работы со стороны любых участников образовательного процесса;

- взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети «Интернет» (электронной почты и т.д.).

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронным ресурсам библиотеки Университета, содержащей издания учебной, учебно-методической и иной литературы по изучаемой дисциплине.

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

В ходе реализации образовательного процесса используются специализированные многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа (практических занятий, лабораторных работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа обеспечивается демонстрационным оборудованием.

Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Для лиц с ограниченными возможностями здоровья и инвалидов предоставляется возможность присутствия в аудитории вместе с ними ассистента (помощника). Для слабовидящих предоставляется возможность увеличения текста на экране ПК. Для самостоятельной работы лиц с ограниченными возможностями здоровья в помещении для самостоятельной работы организовано одно место (ПК) с возможностями бесконтактного ввода информации и управления компьютером (специализированное лицензионное программное обеспечение – Camera Mouse, веб камера). Библиотека университета предоставляет удаленный доступ к электронным ресурсам библиотеки Университета с возможностями для слабовидящих увеличения текста на экране ПК. Лица с ограниченными возможностями здоровья могут при необходимости воспользоваться имеющимся в университете креслом-коляской. В учебном корпусе имеется адаптированный лифт. На первом этаже оборудован специализированный туалет. У входа в здание университета для инвалидов оборудована специальная кнопка, входная среда обеспечена информационной доской о режиме работы университета, выполненной рельефно-точечным тактильным шрифтом (азбука Брайля).