

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 18.08.2025 13:54:37

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51517ff6d591

УТВЕРЖДАЮ:

Ректор

В.В. Волков

«26 » февраля

2025 г.

Протокол УС № 1

от 26.02.2025 г.



Рабочая программа дисциплины
Программирование в R

образовательная программа
направление подготовки
09.04.03 Прикладная информатика

направленность (профиль)
«Прикладной анализ данных и искусственный интеллект»
программа подготовки – магистратура

язык обучения – русский
форма обучения - очная

квалификация (степень) выпускника
Магистр

Санкт-Петербург

Автор:

Котельников Е.В., д. техн. н., доцент, профессор, Школа вычислительных социальных наук, АНООВО «ЕУСПб»

Сосновская Я. А., исследователь, Школа вычислительных социальных наук, АНООВО «ЕУСПб»

Рецензент:

Левшун Д.С., к. тех. н., доцент, Школа вычислительных социальных наук, АНООВО «ЕУСПб»

Рабочая программа дисциплины **«Программирование в R»**, входящей в образовательную программу уровня магистратуры «Прикладной анализ данных и искусственный интеллект», утверждена на заседании Совета Школы вычислительных социальных наук.

Протокол заседания № 4 от 25.02.2025 года.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ **«Программирование в R»**

Дисциплина «Программирование в R» является дисциплиной обязательной части основной профессиональной образовательной программы высшего образования «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика.

Дисциплина «Программирование в R» посвящена изучению основ программирования на языке R с упором на последующую сферу анализа данных. Слушатели изучают основные способы обработки данных в R, инструменты data.table и dplyr, способы визуализации данных в R, основные принципы tidy data, инструменты статистических методов в программной среде R-Studio, создание документов при помощи R Markdown.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточный контроль в форме экзамена.

Общая трудоемкость освоения дисциплины составляет 6 зачетных единиц, 216 часов.

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	8
6.1 Общие положения	8
6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины	8
6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине	9
6.4 Перечень литературы для самостоятельной работы обучающегося	9
6.5 Перечень учебно-методического обеспечения для самостоятельной работы	10
7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ	10
7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации	10
7.2 Контрольные задания для текущей аттестации	11
7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации	12
7.4 Типовые задания к промежуточной аттестации	13
7.5 Средства оценки индикаторов достижения компетенций	16
8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	17
8.1. Основная литература	17
8.2. Дополнительная литература	17
9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	17
9.1 Программное обеспечение	17
9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:	17
9.3 Лицензионные электронные ресурсы библиотеки Университета	18
9.4 Электронная информационно-образовательная среда Университета	18
10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	19
ПРИЛОЖЕНИЕ 1	20

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

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

Задачи:

1. Показать слушателям преимущества языка программирования R и базовую специфику работы с ним.
2. Научить работать с тремя основными парадигмами программирования в R: base-R, dplyr, data.table.
3. Обучить особенностям работы с циклами в R и решении задач через векторизацию.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

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

Планируемые результаты освоения дисциплины, соотнесенные с индикаторами достижения компетенций обучающихся

Таблица 1

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ИД.ОПК-1.1. Решает нестандартные задачи профессиональной деятельности с применением математических, естественнонаучных, социально-экономических методов ИД.ОПК-1.2. В условиях неопределенности и риска способен выработать эффективную стратегию действий, используя математические, естественнонаучные, социально-экономические методы науки	Знать: математические, естественнонаучные и социально-экономические методы для использования в профессиональной деятельности З (ОПК-1) Уметь: решать нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных социально-экономических и профессиональных знаний У (ОПК-1) Владеть: навыком применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач В (ОПК-1)
ОПК-4 Способен применять на практике новые научные принципы и методы исследований	ИД.ОПК-4.1. На основе современных теорий и концепций обосновывает актуальность постановки целей и задач научных исследований в профессиональной области знаний ИД.ОПК-4.2. Анализирует новые научные принципы и методы исследований в профессиональной области знаний ИД.ОПК-4.3. Применяет новые научные принципы и методы	Знать: актуальные направления применения новых научных принципов и методов исследований в профессиональной деятельности З (ОПК-4) Уметь: самостоятельно формировать планы и программы научных исследований с применением новых принципов и методов, характерных для выбранной отрасли науки У (ОПК-4)

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
	исследований в профессиональной области знаний ИД.ОПК-4.4. Разрабатывает предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний	Владеть: навыками системного использования различных новых научных принципов и методов исследований для различных направлений науки В (ОПК-4)

В результате освоения дисциплины магистрант должен:

знатъ:

- Основы синтаксиса языка R, типы данных и структуры.
- Принципы работы с парадигмами программирования в R: base-R, data.table, dplyr.
- Основные методы предобработки и визуализации данных.
- Подходы к анализу данных с использованием статистических методов в R.
- Основы работы с R Markdown и создание воспроизводимых аналитических отчетов.
- Основные методы кластеризации и линейного моделирования в среде R.

уметь:

- Писать чистый и воспроизводимый код на языке R.
- Выполнять базовую и продвинутую предобработку данных с использованием dplyr и data.table.
- Строить графики и визуализировать данные с использованием ggplot2.
- Реализовывать линейную регрессию и интерпретировать ее результаты.
- Применять методы кластеризации к реальным наборам данных.
- Использовать R для подготовки отчетов и презентации результатов анализа.

владеть:

- Навыками программирования на языке R в различных парадигмах.
- Инструментами для анализа и визуализации данных.
- Методами статистического анализа и построения моделей в R.
- Приемами эффективной работы в среде RStudio.
- Техниками отладки, профилирования и документирования кода.
- Навыками воспроизводимого анализа с использованием R Markdown.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Программирование в R» является обязательной дисциплиной Блока 1 «Дисциплины (модули)» обязательной части образовательной программы «Прикладной анализ данных и искусственный интеллект». Курс читается в третьем модуле, форма промежуточной аттестации – экзамен.

Для успешного освоения данной дисциплины требуются знания, полученные в рамках прохождения обучения на уровне бакалавриата/ специалитета.

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе прохождения Б2.О.01(У) Технологической (проектно-технологической) практики и выполнения и защиты выпускной квалификационной работы.

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 6 (шесть) зачетных единиц, 216 часов.

Таблица 2

Объем дисциплины

Типы учебных занятий и самостоятельная работа	Всего	Объем дисциплины									
		Модуль									
		1	2	3	4	5	6	7	8	9	10
Контактная работа обучающихся с преподавателем в соответствии с УП:	28	-	-	28	-	-	-	-	-	-	-
Лекции (Л)	14	-	-	14	-	-	-	-	-	-	-
Лабораторные занятия (ЛЗ)	14	-	-	14	-	-	-	-	-	-	-
Самостоятельная работа (СР)	179	-	-	179	-	-	-	-	-	-	-
Промежуточная аттестация	форма	Экзамен	-	-	Экзамен	-	-	-	-	-	-
	час.	9	-	-	9	-	-	-	-	-	-
Общая трудоемкость дисциплины (час./з.е.)	216/6	-	-	216/6	-	-	-	-	-	-	-

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 соотс Таблицей 1)	Коды ЗУВ (в соотс Таблицей 1)
1	Основы синтаксиса R	Переменные и типы данных, операторы и выражения, структуры данных: векторы, списки, матрицы, датафреймы, функции и области видимости, управление потоком: условия и циклы, основы отладки и тестирования кода, работа с пакетами и библиотеками, ввод и вывод данных	ОПК-1 ОПК-4	ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	3 (ОПК-1) У (ОПК-1) В (ОПК-1) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)
2	R для статистики и анализа данных	Основы использования dplyr и data.table, фильтрация и агрегация данных, объединение таблиц, работа с пропущенными значениями, преобразование текстовых переменных, основы визуализации в ggplot2, построение и визуализация линейной регрессии, методы кластеризации: K-means, DBSCAN, интерпретация результатов	ОПК-1 ОПК-4	ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	3 (ОПК-1) У (ОПК-1) В (ОПК-1) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соотв. с Таблицей 1)	Коды ЗУВ (в соотв. с Таблицей 1)
		статистического анализа, создание отчетов в R Markdown			

5.2 Структура дисциплины

Таблица 4

№ 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	
		Л	ЛЗ		
Очная форма обучения					
Тема 1	Основы синтаксиса R	104	7	7	90
Тема 2	R для статистики и анализа данных	103	7	7	89
Промежуточная аттестация		9	-	-	Экзамен
Всего:		216/6	14	14	179
9					

*Примечание: формы текущего контроля успеваемости: практическое задание (ПЗ).

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

6.1 Общие положения

Знания и навыки, полученные в результате лекций и лабораторных занятий, 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 соответственным образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/лабораторным занятиям. Литературу, рекомендованную в программе курса, следует, по возможности, читать в течение всего семестра, 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 лабораторным занятиям, текущей и промежуточной аттестации по дисциплине также 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. Основы синтаксиса R:

1.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 45 часов.

1.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 45 часов. Итого: 90 часов.

Тема 2. R для статистики и анализа данных:

2.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 40 часов.

2.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 49 часов. Итого: 89 часов.

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

Вопросы для самостоятельной подготовки по темам дисциплины:

1. Какие существуют типы данных в R и как они преобразуются друг в друга?
2. В чем особенности работы со структурами данных: векторы, списки, dataфреймы?
3. Как использовать dplyr для фильтрации, агрегирования и преобразования данных?
4. Какие преимущества предоставляет data.table по сравнению с dplyr?
5. Как выполняется объединение и джойны таблиц в R?
6. Какие функции используются для работы с текстовыми переменными и строками?
7. Как управлять пропущенными значениями в dataфреймах?
8. В чем заключается философия tidy data и как ее реализовать на практике?
9. Как построить график в ggplot2 и какие существуют типы графиков?
10. Какие статистические функции встроены в R и как их применять?
11. Как построить и интерпретировать линейную регрессию в R?
12. Как применить методы кластеризации (K-Means, DBSCAN) в R?
13. В чем особенности написания собственных функций в R?
14. Как использовать R Markdown для автоматизации аналитических отчетов?
15. Какие инструменты RStudio облегчают разработку и отладку кода?

6.4 Перечень литературы для самостоятельной работы обучающегося:

1. Барский, А. Б. Планирование виртуальных вычислений: учеб. пособие / А.Б. Барский. М.: ИД «ФОРУМ»: ИНФРА-М, 2018. 200 с. (Высшее образование). ISBN 978-5-8199-0655-2. Текст: электронный. URL: <https://znanium.com/catalog/product/966062>. Режим доступа: по подписке.

2. 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.

3. Лемешко Б.Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Лемешко Б.Ю., Лемешко С.Б., Постовалов С.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.

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «Программирование 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 лабораторному занятию, выполнение контрольных работ, активное слушание на лекциях. Магистрант должен присутствовать на занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме оценивания выполненных контрольных работ, демонстрирующих степень знакомства 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
Основы синтаксиса R	ОПК-1 ОПК-4	ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Контрольная работа 1	зачтено/ не зачтено
R для статистики и анализа данных	ОПК-1 ОПК-4	ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Контрольная работа 2	зачтено/ не зачтено

Таблица 6

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Контрольная работа	магистрант выполняет задания контрольной работы 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 заданий контрольной работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

7.2 Контрольные задания для текущей аттестации**Примерный материал контрольных работ:****Тема 1: Основы синтаксиса R**

1. Объявите переменные всех основных типов и выведите их на экран.
2. Напишите функцию, возвращающую сумму квадратов элементов вектора.
3. Преобразуйте список в dataфрейм и обратно.
4. Напишите цикл for, суммирующий элементы вектора.
5. Реализуйте условную конструкцию, проверяющую значение переменной.
6. Создайте матрицу 3×3 и выполните транспонирование.
7. Импортируйте CSV-файл и выведите его структуру.
8. Установите пакет ggplot2 и загрузите его.
9. Напишите рекурсивную функцию вычисления факториала.
10. Создайте список с элементами разных типов и получите доступ к каждому.
11. Используйте apply для суммирования строк матрицы.
12. В чем отличие между NA, NULL и NaN в R?
13. Объясните область видимости переменных в функции.
14. Напишите функцию, возвращающую TRUE, если число четное.
15. Создайте свой пакет (структуру пакета) и опишите его назначение.

Тема 2: R для статистики и анализа данных

1. Используя dplyr, выберите строки из dataфрейма по условию.
2. Напишите цепочку операций dplyr для вычисления среднего по группам.
3. Сравните производительность data.table и dplyr на большом наборе данных.
4. Выполните объединение двух таблиц по ключевому столбцу.
5. Напишите код, заменяющий все пропущенные значения на медиану.
6. Используйте stringr для извлечения подстроки из текстового столбца.
7. Постройте график зависимости двух переменных с помощью ggplot2.
8. Визуализируйте распределение категориальной переменной.
9. Постройте линейную регрессию и выведите коэффициенты модели.
10. Постройте диаграмму рассеяния с наложенной регрессионной линией.
11. Примените алгоритм K-means к набору данных iris.
12. Постройте дендрограмму и выполните иерархическую кластеризацию.
13. Создайте отчет в R Markdown, включив код, графики и текст.
14. Выполните анализ выбросов с помощью графических методов.
15. Преобразуйте dataфрейм в формат tidy, используя pivot_longer.

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

Форма промежуточной аттестации – экзамен, выставляемый на основе тестирования.

Перед экзаменом проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Тест включает 20 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балл; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Задания открытого типа

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left(\frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

k_n – максимально возможное количество баллов за вопросы по компетенции,

x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Таблица 7

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)	Критерии оценивания	Оценка
Экзамен / тест	ОПК-1 ОПК-4	ИД.ОПК-1.1.	3 (ОПК-1)	81-100% правильных ответов	Зачлено, отлично
		ИД.ОПК-1.2.	У (ОПК-1)	61-80% правильных ответов	Зачлено, хорошо
		ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	В (ОПК-1) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)	41-60% правильных ответов	Зачлено, удовлетворительно

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соотв. с Таблицей 1)	Коды ЗУВ (в соотв. с Таблицей 1)	Критерии оценивания	Оценка
				0-40% правильных ответов	Не зачленено, неудовлетворительно

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются по стобалльной системе оценки в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 7а.

Таблица 7а
Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачленено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачленено

Результаты промежуточного контроля по дисциплине, выраженные в оценках «отлично», «хорошо», «удовлетворительно», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценке «неудовлетворительно», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

7.4 Типовые задания к промежуточной аттестации

ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте

Комбинированные задания

Повышенный уровень сложности

Задание 1

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.

Вопрос: Из представленных вариантов наиболее подходящим описанием причины популярности языка R является:

Варианты ответа:

- 1) Быстрое выполнение программ за счёт компиляции
- 2) Закрытый исходный код и платное распространение
- 3) Частые обновления, открытый код и активное сообщество
- 4) Удобный графический интерфейс и строгое оформление кода
- 5) Только совместимость с Windows

Поле для ответа:

--	--	--

Обоснование _____

Задание 2

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.

Вопрос: Из представленных вариантов наиболее точно отражает способ выполнения программ в R:

Варианты ответа:

- 1) Программы компилируются в исполняемый файл
- 2) Выполнение происходит строго построчно без возможности изменения
- 3) Интерпретатор выполняет команды интерактивно
- 4) Требуется заранее задать все параметры и переменные
- 5) Код сначала собирается компоновщиком, затем запускается

Поле для ответа:

--	--	--

Обоснование _____

Задание 3

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.

Вопрос: Из представленных утверждений наиболее точно отражает одно из преимуществ R при работе с данными:

Варианты ответа:

- 1) R не поддерживает работу с большими массивами данных
- 2) В R нельзя сохранять результаты анализа
- 3) R позволяет обрабатывать данные разных форматов и сохранять результаты
- 4) R работает только с текстовыми файлами
- 5) В R нельзя подключать внешние библиотеки

Поле для ответа:

--	--	--

Обоснование _____

Задание 4

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.

Вопрос: Наиболее точным описанием графических возможностей R является следующее утверждение:

Варианты ответа:

- 1) R не поддерживает графики
- 2) В R можно строить только гистограммы
- 3) R предоставляет разнообразные инструменты для визуализации данных
- 4) Графики в R создаются только вручную, без автоматизации

5) Для построения графиков в R нужен внешний софт

Поле для ответа:

--	--	--

Обоснование _____

Задание 5

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.

Вопрос: Из следующих утверждений наиболее точно отражает универсальность языка R:

Варианты ответа:

- 1) R используется только в экономике и финансах
- 2) Язык R применяется только для обучения студентов
- 3) R подходит для анализа данных в разных сферах: от социологии до производства
- 4) R можно использовать только в комбинации с Python
- 5) Программы на R работают только в Windows

Поле для ответа:

--	--	--

Обоснование _____

ОПК-4 Способен применять на практике новые научные принципы и методы исследований

Задания открытого типа

Повышенный уровень сложности

Задание 1

Инструкция: Прочтайте задание и запишите развернутый обоснованный ответ

Вопрос: Назовите не менее двух причин, по которым язык программирования R считается популярным инструментом для анализа данных.

Поле для ответа:

Задание 2

Инструкция: Прочтайте задание и запишите развернутый обоснованный ответ

Вопрос: Объясните, в чем состоит преимущество использования интерпретатора в языке R при выполнении анализа данных.

Поле для ответа:

Задание 3

Инструкция: Прочтайте задание и запишите развернутый обоснованный ответ

Вопрос: Укажите два или более способа, с помощью которых R обеспечивает работу с данными различных форматов и объемов.

Поле для ответа:

Задание 4

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Перечислите не менее двух особенностей графических возможностей

языка R, которые делают его удобным для анализа данных.

Поле для ответа: _____

Задание 5

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Объясните, зачем пользователи R создают собственные пакеты и скрипты.

Укажите не менее двух причин.

Поле для ответа: _____

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

Средства оценки индикаторов достижения компетенций

Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Средства оценки (в соот. с Таблицами 5, 7)
ОПК-1	ИД.ОПК-1.1. ИД.ОПК-1.2.	Контрольная работа, тест
ОПК-4	ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	Контрольная работа, тест

Таблица 9

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
Контрольная работа	Магистрант в ходе подготовки и выполнения контрольной работы показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий — На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
Тест	Магистрант в ходе подготовки и выполнения тестирования показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий — На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний

8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

8.1. Основная литература

1. 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.
2. Золотарюк А. В. Язык и среда программирования R: учебное пособие / А.В. Золотарюк. Москва: ИНФРА-М, 2023. 162 с. ISBN 978-5-16-018723-5. Текст: электронный. URL: <https://znanium.com/catalog/product/2049696> . Режим доступа: по подписке.

8.2 Дополнительная литература

1. Барский А. Б. Планирование виртуальных вычислений: учеб. пособие / А.Б. Барский. М.: ИД «ФОРУМ»: ИНФРА-М, 2018. 200 с. (Высшее образование). ISBN 978-5-8199-0655-2. Текст: электронный. URL: <https://znanium.com/catalog/product/966062> . Режим доступа: по подписке.

2. Лемешко Б.Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Лемешко Б.Ю., Лемешко С.Б., Постовалов С.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.

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. ABBYY FineReader 11 Corporate Edition
2. ABBYY Lingvo x5
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. Adobe Acrobat Reader – бесплатно
6. Git (версия 2.40 и выше)
7. Google Chrome
8. Mozilla – бесплатно
9. MS Office (OVS Office Platform)
10. Opera – бесплатно
11. OS Microsoft Windows (OVS OS Platform)
12. VLC – бесплатно
13. Яндекс.Браузер (Yandex Browser) – бесплатно
14. R – 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://proed.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. ЕНИП — Электронная библиотека «Научное наследие России»: <http://e-heritage.ru/>

2. Интелрос. Интеллектуальная Россия: <http://www.intelros.ru/>

3. Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>

4. Президентская библиотека: <http://www.prlib.ru>

5. Российская государственная библиотека: <http://www.rsl.ru/>

6. Российская национальная библиотека: <http://www.nlr.ru/poisk/>

9.3 Лицензионные электронные ресурсы библиотеки Университета

Профессиональные базы данных:

Полный перечень доступных обучающимся профессиональных баз данных представлен на официальном сайте Университета <https://eusp.org/library/electronic-resources>, включая следующие базы данных:

1. **eLIBRARY.RU** — Российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, научометрическая база данных: <http://elibrary.ru>;

2. Электронные журналы по подписке (текущие номера научных зарубежных журналов)

Электронные библиотечные системы:

1. **Znanium.com** — Электронная библиотечная система (ЭБС) — <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://eusp.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).

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ
«Программирование в R»

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

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 лабораторному занятию, выполнение контрольных работ, активное слушание на лекциях. Магистрант должен присутствовать на занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме оценивания выполненных контрольных работ, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 1

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
Основы синтаксиса R	ОПК-1 ОПК-4	ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Контрольная работа 1	зачтено/ не зачтено
R для статистики и анализа данных	ОПК-1 ОПК-4	ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Контрольная работа 2	зачтено/ не зачтено

Таблица 2

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Контрольная работа	магистрант выполняет задания контрольной работы 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 заданий контрольной работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

2 Контрольные задания для текущей аттестации

Материал контрольных работ:

Тема 1: Основы синтаксиса R

1. Объявите переменные всех основных типов и выведите их на экран.
2. Напишите функцию, возвращающую сумму квадратов элементов вектора.

3. Преобразуйте список в датафрейм и обратно.
4. Напишите цикл `for`, суммирующий элементы вектора.
5. Реализуйте условную конструкцию, проверяющую значение переменной.
6. Создайте матрицу 3×3 и выполните транспонирование.
7. Импортируйте CSV-файл и выведите его структуру.
8. Установите пакет `ggplot2` и загрузите его.
9. Напишите рекурсивную функцию вычисления факториала.
10. Создайте список с элементами разных типов и получите доступ к каждому.
11. Используйте `apply` для суммирования строк матрицы.
12. В чем отличие между `NA`, `NULL` и `NaN` в R?
13. Объясните область видимости переменных в функции.
14. Напишите функцию, возвращающую `TRUE`, если число четное.
15. Создайте свой пакет (структуру пакета) и опишите его назначение.

Тема 2: R для статистики и анализа данных

1. Используя `dplyr`, выберите строки из датафрейма по условию.
2. Напишите цепочку операций `dplyr` для вычисления среднего по группам.
3. Сравните производительность `data.table` и `dplyr` на большом наборе данных.
4. Выполните объединение двух таблиц по ключевому столбцу.
5. Напишите код, заменяющий все пропущенные значения на медиану.
6. Используйте `stringr` для извлечения подстроки из текстового столбца.
7. Постройте график зависимости двух переменных с помощью `ggplot2`.
8. Визуализируйте распределение категориальной переменной.
9. Постройте линейную регрессию и выведите коэффициенты модели.
10. Постройте диаграмму рассеяния с наложенной регрессионной линией.
11. Примените алгоритм K-means к набору данных `iris`.
12. Постройте дендрограмму и выполните иерархическую кластеризацию.
13. Создайте отчет в R Markdown, включив код, графики и текст.
14. Выполните анализ выбросов с помощью графических методов.
15. Преобразуйте датафрейм в формат `tidy`, используя `pivot_longer`.

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

Форма промежуточной аттестации – экзамен, выставляемый на основе тестирования.

Перед экзаменом проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Тест включает 20 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балл; если на любой одной позиции ответа

записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Задания открытого типа

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left(\frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,
 K – количество осваиваемых в рамках дисциплины компетенций,
 k_n – максимально возможное количество баллов за вопросы по компетенции,
 x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Таблица 3
Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соотв. с Таблицей 1)	Коды ЗУВ (в соотв. с Таблицей 1)	Критерии оценивания	Оценка
Экзамен / тест	ОПК-1 ОПК-4	ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	3 (ОПК-1) У (ОПК-1) В (ОПК-1) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)	81-100% правильных ответов	Зачтено, отлично
				61-80% правильных ответов	Зачтено, хорошо
				41-60% правильных ответов	Зачтено, удовлетворительно
				0-40% правильных ответов	Не зачтено, неудовлетворительно

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются по стобалльной системе оценки в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 3а.

Таблица 3а
Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в оценках «отлично», «хорошо», «удовлетворительно», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций

образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценке «неудовлетворительно», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

4 Задания к промежуточной аттестации

ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте

Базовый уровень сложности

1. Прочтайте задание, выберите правильный ответ

В R индексация векторов начинается с:

1. 0
2. 1
3. -1
4. Зависит от типа данных

Поле для ответа:

2. Прочтайте задание, выберите правильный ответ

Какой оператор в R позволяет проверить, содержится ли значение в векторе?

1. ==
2. %in%
3. &
4. |

Поле для ответа:

3. Прочтайте задание, выберите правильный ответ

Какие из перечисленных операторов и функций используются в R для работы с текстовыми данными?

1. +
2. %in%
3. gsub()
4. grep()

Поле для ответа:

4. Прочтайте задание, выберите правильный ответ

Какое из следующих утверждений НЕ верно относительно типов данных в R?

1. Логические значения записываются как TRUE и FALSE
2. Целые числа автоматически получают тип integer
3. Факторы используются для работы с категориальными данными
4. Дата и время хранятся в специальном формате POSIXct

Поле для ответа:

5. Прочтайте задание, выберите правильный ответ

Какая функция для фильтрации строк датафрейма по условию в пакете dplyr позволяет оставить только те наблюдения, которые соответствуют заданному критерию?

1. select() – выбор столбцов по названию
2. filter() – отбор строк по условию
3. mutate() – создание новых переменных
4. summarize() – агрегация данных

Поле для ответа:

6. Прочтайте задание, выберите правильный ответ

В R для объединения двух датафреймов по общему столбцу в пакете dplyr используется семейство функций join. Какая из них оставляет только те строки, где есть совпадения в обоих датафреймах?

1. left_join() – сохраняет все строки из левого датафрейма
2. right_join() – сохраняет все строки из правого датафрейма
3. inner_join() – оставляет только совпадающие строки
4. full_join() – сохраняет все строки из обоих датафреймов

Поле для ответа:

7. Прочтайте задание, выберите правильный ответ

При работе с датами в R иногда требуется преобразовать строку в формат POSIXct или Date. Какие функции позволяют это сделать?

1. as.numeric() – преобразование в число

2. `as.character()` – преобразование в строку
3. `as.POSIXct()` – преобразование в дату и время
4. `as.Date()` – преобразование в дату

Поле для ответа:

8. Прочтайте задание, выберите правильный ответ

При обработке данных в R часто требуется фильтрация. Какие из этих выражений правильно фильтруют данные по условию?

1. `df[df$age > 18,]`
2. `df[, df$gender == "male"]`
3. `subset(df, salary > 50000)`
4. `df[df$score]`

Поле для ответа:

9. Прочтайте задание, выберите правильный ответ

В R для объединения датафреймов по общему столбцу используются функции семейства `join`. Какие из перечисленных функций сохраняют все строки из хотя бы одного исходного датафрейма?

1. `inner_join()` – только совпадающие строки
2. `left_join()` – все строки левого датафрейма
3. `right_join()` – все строки правого датафрейма
4. `anti_join()` – строки без совпадений

Поле для ответа:

10. Прочтайте задание, выберите правильный ответ

При обработке текстовых данных часто требуется удалить стоп-слова (например, предлоги и местоимения). Из представленных вариантов наиболее эффективным способом сделать это в R является:

1. Вручную создать список стоп-слов и фильтровать через `filter()`
2. Использовать функцию `get_stopwords()` из пакета `tidytext`
3. Применить `str_replace_all()` для удаления отдельных слов
4. Использовать `str_detect()` для поиска стоп-слов

Поле для ответа:

11. Прочтите задание, выберите правильный ответ

Для токенизации текста (разбиения на отдельные слова) в R чаще всего используют:

1. Функцию `str_split()` из базового R
2. Функцию `unnest_tokens()` из пакета `tidytext`
3. Функцию `stri_split()` из пакета `stringi`
4. Функцию `separate()` из `tidyverse`

Поле для ответа:

12. Прочтите задание, выберите правильный ответ

В R для работы со строками используются разные пакеты. Какие функции из перечисленных относятся к пакету `stringr`?

1. `stri_length()`
2. `str_detect()`
3. `gsub()`
4. `str_replace()`

Поле для ответа:

13. Прочтите задание, выберите правильный ответ

Какие из перечисленных функций используются для соединения строк в R?

1. `str_split()`
2. `str_c()`
3. `paste()`
4. `str_detect()`

Поле для ответа:

14. Прочтите задание, выберите правильный ответ

Какие из перечисленных функций позволяют найти подстроку в тексте?

1. `str_detect()`
2. `str_count()`
3. `str_replace()`
4. `str_sort()`

15. Прочтите задание, выберите правильный ответ

Если на веб-странице нужный элемент не имеет класса или ID, но его можно найти по структуре HTML-дерева, какой метод извлечения данных следует использовать?

1. CSS-селекторы
2. XPath
3. Регулярные выражения
4. JSON-парсинг

Поле для ответа:

Повышенный уровень сложности

1. Прочтайте задание и установите последовательность.

Для корректной работы с датафреймом в R необходимо выполнить ряд действий. Установите правильную последовательность операций при создании и обработке датафрейма:

1. Присвоить имена колонкам
2. Создать векторы с данными
3. Объединить векторы в датафрейм
4. Фильтровать данные по условию
5. Добавить новую колонку

Поле для ответа:

2. Прочтайте задание и установите последовательность.

При загрузке данных из файла в R важно соблюдать правильный порядок действий. Установите последовательность:

1. Указать путь к файлу
2. Проверить кодировку данных
3. Загрузить данные функцией `read.csv()`
4. Установить рабочую директорию
5. Просмотреть структуру данных

Поле для ответа:

3. Прочтайте задание и установите последовательность.

Для создания матрицы в R требуется выполнить несколько шагов. Установите правильную последовательность:

1. Задать имена строк и столбцов
2. Указать количество строк и столбцов
3. Создать вектор с данными
4. Преобразовать вектор в матрицу

5. Выполнить операции с матрицей

Поле для ответа:

4. Прочтайте задание и установите последовательность.

При работе с библиотеками в R необходимо следовать определённому порядку.

Установите последовательность действий:

1. Установить библиотеку
2. Проверить доступные функции
3. Активировать библиотеку
4. Импортировать данные
5. Использовать функции библиотеки

Поле для ответа:

5. Прочтайте задание и установите последовательность.

Для обработки текстовых данных в R выполняется ряд операций. Установите правильную последовательность:

1. Заменить текст функцией gsub()
2. Создать текстовый вектор
3. Проверить наличие подстроки grep()
4. Разделить текст на части
5. Преобразовать регистр символов

Поле для ответа:

6. Прочтайте задание и установите последовательность.

При сохранении результатов анализа в R важно соблюдать порядок действий. Установите последовательность:

1. Указать имя файла
2. Выбрать формат сохранения
3. Проверить данные перед сохранением
4. Задать рабочую директорию
5. Выполнить команду сохранения

Поле для ответа:

7. Прочтайте задание и установите последовательность.

В R для работы с данными часто используется пакет `dplyr`. Чтобы корректно загрузить и использовать его функции, необходимо выполнить ряд действий. Установите правильную последовательность операций при работе с библиотеками в R.

1. Установить пакет с помощью `install.packages('dplyr')`
2. Вызвать функцию напрямую через `dplyr::select()`
3. Загрузить библиотеку в сессию с помощью `library(dplyr)`
4. Проверить конфликт имён функций
5. Прочитать документацию по пакету

Поле для ответа:

8. Прочтайте задание и установите последовательность.

При загрузке данных в R из CSV-файла могут возникнуть проблемы с кодировкой. Укажите правильный порядок действий для корректного чтения файла с кириллическим текстом.

1. Указать аргумент `header = TRUE`
2. Использовать функцию `read.csv()`
3. Добавить параметр `encoding = "UTF-8"`
4. Проверить данные на корректность
5. Указать путь к файлу

Поле для ответа:

9. Прочтайте задание и установите последовательность.

Для анализа данных в `dplyr` часто применяются пайплайны (`%>%`). Установите правильную последовательность операций при обработке датафрейма.

1. Отфильтровать строки с помощью `filter()`
2. Сгруппировать данные по переменной `group_by()`
3. Выбрать нужные столбцы `select()`
4. Посчитать агрегированные значения `summarize()`
5. Отсортировать результаты `arrange()`

Поле для ответа:

10. Прочтайте задание и установите последовательность.

При объединении датафреймов в R важно правильно выбрать тип соединения. Укажите последовательность действий для корректного слияния таблиц по общему столбцу.

1. Указать общий ключ в `join_by()`
2. Выбрать тип соединения (`inner_join`, `left_join` и др.)

3. Загрузить оба датафрейма
4. Проверить результат на пропуски
5. Применить функцию слияния

Поле для ответа:

11. Прочитайте задание и установите последовательность.

Для агрегации данных в `dplyr` используется комбинация `group_by()` и `summarize()`. Укажите правильный порядок действий для расчёта среднего значения по группам.

1. Задать функцию агрегации (`mean`, `sum` и др.)
2. Сгруппировать данные по нужному столбцу
3. Применить `summarize()`
4. Вывести результат
5. Отфильтровать данные (если нужно)

Поле для ответа:

12. Прочитайте задание и установите последовательность.

При работе с датами в R иногда требуется преобразовать формат `POSIXct` в более простой. Укажите последовательность действий для конвертации времени в дату.

1. Загрузить данные с временными метками
2. Использовать функцию `as.Date()`
3. Проверить новый формат
4. Убедиться, что столбец имеет тип `POSIXct`
5. Применить преобразование

Поле для ответа:

13. Прочитайте задание и установите последовательность.

Для преобразования данных из широкого формата в длинный в R используется функция `melt()`. Установите последовательность действий, чтобы корректно выполнить это преобразование.

1. Указать колонки, которые останутся без изменений (`id.vars`)
2. Загрузить пакет `reshape2`
3. Выбрать колонки, которые нужно преобразовать (`measure.vars`)
4. Применить функцию `melt()` к датафрейму
5. Задать имена для новой колонки с категориями (`variable.name`)

Поле для ответа:

14. Прочтите задание и установите последовательность.

При работе с грязными данными важно провести их очистку. Установите последовательность действий, которые необходимо выполнить, чтобы преобразовать данные с запятыми в десятичных числах в корректный числовой формат.

1. Заменить запятые на точки в числовых колонках
2. Проверить структуру данных (`str()`)
3. Преобразовать колонки в тип `numeric`
4. Удалить лишние пробелы в текстовых полях
5. Отфильтровать строки с пропущенными значениями

Поле для ответа:

15. Прочтите задание и установите последовательность.

Для визуализации данных в RMarkdown с помощью `kableExtra` нужно выполнить несколько действий. Какая последовательность приведет к созданию стилизованной таблицы?

1. Применить `kable_classic()` для базового оформления
2. Загрузить данные (`data("USArrests")`)
3. Использовать `kable()` для создания таблицы
4. Добавить заголовок (`caption`)
5. Выделить строки с помощью `row_spec()`

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

1. Прочтайте текст задания и запишите развернутый обоснованный ответ.

Какой тип данных в R предназначен для хранения табличных данных?

Поле для ответа:

2. Прочтайте текст задания и запишите развернутый обоснованный ответ.

В каких типах объекта могут храниться только однородные данные?

Поле для ответа:

3. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какой тип объекта в R используется для хранения разнородных данных (например, чисел, текста и других структур)?

Поле для ответа:

4. Прочтите текст задания и запишите развернутый обоснованный ответ.

`vector[0]` - корректно ли такое обращение к первому элементу вектора?

Поле для ответа:

5. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какие из этих операций можно выполнять с матрицами в R?

Поле для ответа:

6. Прочтите текст задания и запишите развернутый обоснованный ответ.

При загрузке данных из CSV-файла в R может возникнуть проблема с кодировкой, из-за которой текст отображается некорректно. Какой аргумент функции `read.csv()` позволяет исправить эту проблему?

Поле для ответа:

7. Прочтите текст задания и запишите развернутый обоснованный ответ.

В R для обработки текстовых данных часто используются функции из пакета `stringr`. Какие операции можно выполнить с его помощью?

Поле для ответа:

8. Прочтите текст задания и запишите развернутый обоснованный ответ.

При анализе данных часто требуется преобразовать таблицу из широкого формата в длинный. Какой инструмент в R позволяет сделать это наиболее удобно?

Поле для ответа:

9. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какая из проблем чаще всего встречается в "грязных" данных и требует предварительной обработки перед анализом?

Поле для ответа:

10. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какой формат данных лучше подходит для анализа временных рядов и группировки значений?

Поле для ответа:

ОПК-4 Способен применять на практике новые научные принципы и методы исследований

Задания закрытого типа

Базовый уровень сложности

1. Прочтайте задание, выберите правильный ответ

Какой тип графика позволяет одновременно отобразить медиану, квартили и выбросы в данных?

1. Диаграмма плотности
2. Столбчатая диаграмма
3. Ящик с усами (Box plot)
4. Линейчатая диаграмма

Поле для ответа:

2. Прочтайте задание, выберите правильный ответ

Какой параметр в `ggplot2` отвечает за разбиение графика на подграфики по категориальной переменной?

1. `aes()`
2. `facet_grid()`
3. `theme()`
4. `geom_point()`

Поле для ответа:

3. Прочтайте задание, выберите правильный ответ

Какой график лучше всего подходит для визуализации взаимосвязи между двумя числовыми переменными?

1. Гистограмма
2. Диаграмма рассеяния (Scatter plot)
3. Столбчатая диаграмма
4. Box plot

Поле для ответа:

4. Прочтайте задание, выберите правильный ответ

Какой параметр в `ggplot2` позволяет добавить линейный тренд на точечный график?

1. `geom_line()`
2. `geom_smooth()`
3. `geom_bar()`
4. `geom_histogram()`

Поле для ответа:

5. Прочтайте задание, выберите правильный ответ

В описательной статистике используются различные меры для анализа данных. Какие из перечисленных показателей относятся к мерам центральной тенденции?

1. Среднее арифметическое
2. Дисперсия
3. Медиана
4. Стандартное отклонение
5. Мода
6. Квантиль

Поле для ответа:

6. Прочтайте задание, выберите правильный ответ

В `ggplot2` для построения графиков используются слои. Какие из перечисленных функций относятся к геометрическим слоям (`geom_*`)?

1. `geom_point()`
2. `aes()`
3. `geom_histogram()`
4. `theme_minimal()`
5. `geom_boxplot()`

6. `ggttitle()`

Поле для ответа:

7. Прочтите задание, выберите правильный ответ

В `ggplot2` для добавления дополнительной категориальной переменной на график можно использовать разные эстетики. Какие из перечисленных параметров `aes()` позволяют это сделать?

- 1. `x`
- 2. `color`
- 3. `fill`
- 4. `size`
- 5. `shape`

Поле для ответа:

8. Прочтите задание, выберите правильный ответ

При сохранении графика в `ggplot2` с помощью `ggsave()` можно настроить параметры вывода. Какие из следующих аргументов влияют на качество и размер изображения?

- 1. `filename`
- 2. `plot`
- 3. `width`
- 4. `height`
- 5. `dpi`

Поле для ответа:

9. Прочтите задание, выберите правильный ответ

Если при парсинге веб-страницы нужно извлечь не текст, а ссылку, встроенную в элемент, какую функцию пакета `rvest` следует использовать?

- 1. `html_text()`
- 2. `html_attr()`
- 3. `html_nodes()`
- 4. `html_table()`

Поле для ответа:

10. Прочтите задание, выберите правильный ответ

При парсинге веб-страниц с помощью `rvest` важно правильно указать путь к нужному элементу. Какие из перечисленных способов позволяют корректно извлечь текстовые данные из HTML-разметки?

1. Использование `html_node()` с указанием класса элемента
2. Применение `html_attr()` для извлечения текста
3. Использование `html_text()` после выбора узла
4. Извлечение данных через `html_links()`

Поле для ответа:

11. Прочтите задание, выберите правильный ответ

Какие функции из пакета `rvest` используются для извлечения данных из таблиц на веб-странице?

1. `html_table()`
2. `html_text()`
3. `html_nodes()`
4. `read_csv()`

Поле для ответа:

12. Прочтите задание, выберите правильный ответ

Какой метод позволяет скачать файлы с веб-страницы автоматически?

1. `download_html()`
2. `download.file()`
3. `read_html()`
4. `html_download()`

Поле для ответа:

13. Прочтите задание, выберите правильный ответ

При анализе данных в R иногда требуется фильтровать наблюдения. Какие из следующих операций корректно отфильтруют строки с нулевыми значениями ИМТ (BMI) в датасете `pima`?

1. `pima %>% filter(BMI > 0)`
2. `pima %>% filter(BMI != 0)`
3. `pima %>% filter(!is.na(BMI))`
4. `pima %>% filter(BMI == 0)`
5. `pima %>% filter(BMI >= 0)`

14. Прочтите задание, выберите правильный ответ

В каких случаях применяется тест Хи-квадрат?

1. Для сравнения средних значений в двух группах
2. Для проверки связи между двумя категориальными переменными
3. Для анализа нормальности распределения
4. Для сравнения пропорций в таблицах сопряженности
5. Для оценки корреляции между непрерывными переменными

Поле для ответа:

15. Прочтите задание, выберите правильный ответ

Какие условия должны выполняться перед проведением t-теста?

1. Данные должны быть категориальными
2. Данные должны быть распределены нормально
3. Дисперсии в группах должны быть равны
4. Выборка должна быть очень большой ($n > 1000$)
5. Переменная должна быть непрерывной

Поле для ответа:

Повышенный уровень сложности

1. Прочтите задание и установите последовательность.

Установите последовательность действий для анализа текстовых данных в R.

1. Удалить стоп-слова
2. Загрузить текстовые данные
3. Подсчитать частотность слов
4. Очистить текст от лишних символов
5. Токенизировать текст

Поле для ответа:

2. Прочтите задание и установите последовательность.

Установите последовательность действия для подсчета частотности слов.

1. Отсортировать результаты по убыванию
2. Удалить стоп-слова
3. Отфильтровать короткие слова
4. Подсчитать частоту

Поле для ответа:

3. Прочтайте задание и установите последовательность.

Установите последовательность действий для обработки нескольких страниц пагинации.

1. Организовать внутренний цикл по статьям на странице
2. Определить шаблон URL страниц (например, "/page/")
3. Сохранить данные в общий датафрейм
4. Организовать внешний цикл по страницам

Поле для ответа:

4. Прочтайте задание и установите последовательность.

При обработке динамических элементов страницы может потребоваться использование XPath вместо CSS-селекторов. Установите последовательность действий.

1. Найти элемент в разметке HTML
2. Скопировать XPath через контекстное меню
3. Использовать `html_node(xpath=...)` вместо CSS-селектора
4. Проверить, есть ли у элемента классы

Поле для ответа:

5. Прочтайте задание и установите последовательность.

Установите последовательность действий, необходимых при парсинге нескольких страниц сайта.

1. Создать пустой датафрейм для результатов
2. Добавить задержку между запросами с помощью `Sys.sleep()`
3. Определить шаблон URL страниц
4. В цикле обрабатывать каждую страницу и сохранять данные

Поле для ответа:

6. Прочтайте задание и установите последовательность.

Установите последовательность действий для построения графика в `ggplot2`.

1. Загрузить данные (`pima <- read.csv('diabetes.csv')`)
2. Указать данные и оси (`aes(x = Glucose, y = diabetes_ch)`)
3. Добавить тип графика (`geom_boxplot()`)
4. Настроить тему (`theme_minimal()`)
5. Создать базовый слой (`ggplot()`)

Поле для ответа:

7. Прочтайте задание и установите последовательность.

Установите последовательность действия при анализе данных с помощью описательной статистики.

1. Рассчитать среднее и медиану (`mean(pima$Glucose)`)
2. Построить гистограмму (`geom_histogram(aes(x = Glucose))`)
3. Загрузить данные (`pima <- read.csv('diabetes.csv')`)
4. Проверить выбросы (`boxplot(pima$Glucose)`)
5. Установить библиотеки (`library(ggplot2)`)

Поле для ответа:

8. Прочтайте задание и установите последовательность.

Установите последовательность действий при сохранении графика в файл с помощью `ggsave()`.

1. Указать имя файла ("plot.png")
2. Задать качество (`dpi = 1000`)
3. Построить график (`ggplot() + geom_point()`)
4. Указать размеры (`width = 8.5, height = 9`)
5. Сохранить (`ggsave()`)

Поле для ответа:

9. Прочтайте задание и установите последовательность.

Установите последовательность действий для построения scatter plot с линейным трендом в `ggplot2`.

1. Добавить точки (`geom_point()`)
2. Отфильтровать данные (`filter(BMI != 0)`)
3. Задать линейный тренд (`geom_smooth(method = lm)`)
4. Указать оси (`aes(x = BMI, y = SkinThickness)`)
5. Создать базовый слой (`ggplot()`)

Поле для ответа:

10. Прочтайте задание и установите последовательность.

Установите последовательность действий при создании столбчатой диаграммы с группировкой по категориям.

1. Указать переменную для группировки (aes(fill = diabetes_ch))
2. Добавить столбцы (geom_bar(position = "dodge"))
3. Задать ось X (aes(x = age_group))
4. Настроить тему (theme_minimal())
5. Создать базовый слой (ggplot(data = pima))

Поле для ответа:

11. Прочтайте задание и установите последовательность.

Перед применением параметрических тестов необходимо проверить, нормально ли распределены данные. Для этого используется тест Шапиро-Уилка. Укажите правильную последовательность действий при проверке нормальности распределения.

1. Визуализировать распределение с помощью гистограммы
2. Провести тест Шапиро-Уилка (shapiro.test())
3. Интерпретировать p-value
4. Загрузить данные в R
5. Если p-value < 0.05, распределение не нормальное

Правильный ответ:

12. Прочтайте задание и установите последовательность.

Корреляционный анализ позволяет оценить взаимосвязь между двумя переменными. Выберите верную последовательность действий для расчета корреляции Пирсона в R.

1. Проверить нормальность распределения переменных
2. Загрузить данные (data <- read.csv(...))
3. Построить scatterplot для визуализации связи (ggplot() + geom_point())
4. Рассчитать корреляцию (cor.test(x, y, method = "pearson"))
5. Интерпретировать коэффициент корреляции и p-value

Поле для ответа:

13. Прочтайте задание и установите последовательность.

T-тест используется для сравнения средних значений в двух группах. Укажите правильный порядок действий при проведении t-теста в R.

1. Проверить равенство дисперсий (var.test())
2. Провести t-тест (t.test(group1, group2, var.equal = ...))
3. Загрузить данные и выделить группы
4. Интерпретировать p-value

5. Проверить нормальность распределения (`shapiro.test()`)

Поле для ответа:

14. Прочтите задание и установите последовательность.

Выбор статистического теста зависит от типа данных и целей анализа. Укажите правильную последовательность шагов для выбора теста.

1. Определить, нормально ли распределены данные
2. Определить тип переменных (непрерывные/категориальные)
3. Выбрать параметрический или непараметрический тест
4. Провести тест
5. Интерпретировать результаты

Поле для ответа:

15. Прочтите задание и установите последовательность.

Перед проведением статистического анализа важно визуализировать данные. Выберите правильный порядок действий для предварительного анализа.

1. Построить гистограмму (`hist()`)
2. Построить boxplot (`ggplot() + geom_boxplot()`)
3. Загрузить данные
4. Проверить выбросы
5. Выбрать подходящий статистический тест

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

1. Прочтайте текст задания и запишите развернутый обоснованный ответ.

Какой специализированный пакет чаще всего используют в R для обработки текстовых данных с помощью регулярных выражений?

Поле для ответа:

2. Прочтайте текст задания и запишите развернутый обоснованный ответ.

Какие действия относятся к предварительной обработке текстовых данных перед анализом?

Поле для ответа:

3. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какие из операции выполняются при токенизации текста?

Поле для ответа:

4. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какой инструмент в R является наиболее подходящим для сбора данных с веб-страниц, где контент загружается динамически (например, через JavaScript)?

Поле для ответа:

5. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какая команда R наиболее удобна для скачивания файлов с веб-страницы, если известны их прямые ссылки?

Поле для ответа:

6. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какой слой веб-страницы отвечает за оформление шрифтов и цветов?

Поле для ответа:

7. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какой элемент HTML-разметки чаще всего содержит ссылки на странице?

Поле для ответа:

8. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какой тип графика лучше всего подходит для первичного анализа данных и визуализации распределения числовой переменной?

Поле для ответа:

9. Прочтите текст задания и запишите развернутый обоснованный ответ.

Какие типы графиков подходят для визуализации распределения одной числовой переменной?

Поле для ответа:

10. Прочтите текст задания и запишите развернутый обоснованный ответ.

В каких случаях следует использовать непараметрические тесты вместо параметрических?

Поле для ответа:

5 Средства оценки индикаторов достижения компетенций

Таблица 4

Средства оценки индикаторов достижения компетенций

Коды компетенций	Индикаторы компетенций (в соотв. с Таблицей 1)	Средства оценки (в соотв. с Таблицами 5, 7)
ОПК-1	ИД.ОПК-1.1. ИД.ОПК-1.2.	Контрольная работа, тест
ОПК-4	ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	Контрольная работа, тест

Таблица 5

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соотв. С Таблицами 5, 7)	Рекомендованный план выполнения работы
Контрольная работа	Магистрант в ходе подготовки и выполнения контрольной работы показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий — На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
Тест	Магистрант в ходе подготовки и выполнения тестирования показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности:

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
	<ul style="list-style-type: none"> — Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий — На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний