

Документ подписан простой электронной подписью.
Информация о владельце:

ФИО: Кандрашина Елена Александровна

Должность: И.о. ректора ФГАОУ ВО «Самарский государственный экономический университет»

Дата подписания: 24.09.2025 10:51:56

Уникальный программный ключ:

2db64eb9605ce27edd3b8e8fdd32c70e0674ddd2

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Самарский государственный экономический университет»

Институт Институт экономики предприятий

Кафедра Статистики и эконометрики

УТВЕРЖДЕНО

Ученым советом Университета

(протокол №1 от 29 августа 2025 г.)

РАБОЧАЯ ПРОГРАММА

Наименование дисциплины Б1.О.14 Анализ больших данных

Основная профессиональная образовательная программа 27.03.02 Управление качеством
Экономика и управление качеством

Квалификация (степень) выпускника Бакалавр

Самара 2025

Целью изучения дисциплины является формирование результатов обучения, обеспечивающих достижение планируемых результатов освоения образовательной программы.

1. Место дисциплины в структуре ОП

Дисциплина Анализ больших данных входит в обязательную часть блока Б1.Дисциплины (модули).

Компетенции	Предшествующие дисциплины	Последующие дисциплины
ОПК-7 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности		
ОПК-7.2 Владеет современными информационными технологиями для поиска и обмена информацией при решении задач в области управления качеством	Технологические основы производства Маркетинг	Маркетинг Основы проектной деятельности Инвестиционный анализ Практикум по управлению производственным предприятием Производственная практика: организационно-управленческая практика Производственная практика: преддипломная практика Подготовка к процедуре защиты и защита выпускной квалификационной работы
ПК-5 - Способен проводить внутренний аудит системы управления качеством организации, анализировать его результаты, разрабатывать мероприятия по предотвращению выпуска некачественной продукции		
ПК-5.3 Анализирует и выбирает оптимальные методы и методики решения конкретных задач по управлению качеством продукции (работ, услуг)		Деловые коммуникации и документооборот Управление рисками Системы менеджмента качества Бережливое производство Управление затратами на обеспечение качества Стандарты качества в управленческой деятельности Производственная практика: организационно-управленческая практика Производственная практика: преддипломная практика Производственная практика: практика по профилю профессиональной деятельности Подготовка к процедуре защиты и защита выпускной квалификационной работы

2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе

Изучение дисциплины Анализ больших данных в образовательной программе направлено на формирование у обучающихся следующих компетенций:

Общепрофессиональные компетенции (ОПК):

ОПК-7 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
ОПК-7.2 Владеет современными информационными технологиями для поиска и обмена информацией при решении задач в области управления качеством	ОПК-7.2: Знать современные информационные технологии, используемые при решении задач в области управления качеством	ОПК-7.2: Уметь искать и обмениваться информацией с помощью ИТ при решении задач в области управления качеством	ОПК-7.2: Владеть современными информационными технологиями для поиска и обмена информацией при решении задач в области управления качеством

Профессиональные компетенции (ПК):

ПК-5 - Способен проводить внутренний аудит системы управления качеством организации, анализировать его результаты, разрабатывать мероприятия по предотвращению выпуска некачественной продукции

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
ПК-5.3 Анализирует и выбирает оптимальные методы и методики решения конкретных задач по управлению качеством продукции (работ, услуг)	ПК-5.3: Знать разнообразные методы и методики решения конкретных задач по управлению качеством продукции (работ, услуг)	ПК-5.3: Уметь анализировать и выбирать оптимальные методы и методики решения конкретных задач по управлению качеством продукции (работ, услуг)	ПК-5.3: Владеть навыками анализа оптимальных методов и методик решения конкретных задач по управлению качеством продукции (работ, услуг)

3. Объем и виды учебной работы

Учебным планом предусматриваются следующие виды учебной работы по дисциплине:

Очная форма обучения

Виды учебной работы	Всего час/ з.е.
	Сем 3
Контактная работа, в том числе:	56.3/1.56
Занятия лекционного типа	36/1
Занятия семинарского типа	18/0.5
Индивидуальная контактная работа (ИКР)	0.3/0.01

Групповая контактная работа (ГКР)	2/0.06
Самостоятельная работа:	53.7/1.49
Промежуточная аттестация	34/0.94
Вид промежуточной аттестации: Экзамен	Экз
Общая трудоемкость (объем части образовательной программы): Часы	144
Зачетные единицы	4

4. Содержание дисциплины

Тематический план дисциплины Анализ больших данных представлен в таблице.

Разделы, темы дисциплины и виды занятий Очная форма обучения

№ п/п	Наименование темы (раздела) дисциплины	Контактная работа				Самостоятельная работа	Планируемые результаты обучения в соотношении с результатами обучения по образовательной программе
		Лекции	Занятия семинарского типа	ИКР	ГКР		
			Практич. занятия				
	Раздел 1. Современные инструменты и алгоритмы анализа BIG DATA.						
1.	Введение в Big Data: определение и ключевые концепции (основные характеристики больших данных, их структура и объемы; применение Big Data в различных отраслях; Big Data при анализе результатов деятельности предприятий)	2				3	ОПК-7.2, ПК-5.3
2.	Инструменты для обработки и хранения больших данных (обзор технологий хранения: Hadoop, Spark, NoSQL базы данных (Cassandra, MongoDB); сравнение традиционных реляционных баз данных и NoSQL)	2				3	ОПК-7.2, ПК-5.3
3.	Алгоритмы машинного обучения для анализа больших данных (введение в машинное обучение: supervised и unsupervised методы; применение таких алгоритмов как деревья решений и кластеризация к большим данным; использование машинного обучения при анализе результатов деятельности предприятий)	2				3	ОПК-7.2, ПК-5.3
4.	Обработка потоковых данных (основы потоковой обработки Apache Kafka,	2				3	ОПК-7.2, ПК-5.3

	Apache Flink; использование потоковых данных в реальном времени на примерах)						
5.	Технологии визуализации данных для больших объемов информации (инструменты визуализации: Tableau, Power BI, D3.js; принципы эффективной визуализации больших данных)	2				3	ОПК-7.2, ПК-5.3
6.	Анализ текстовых данных и обработка естественного языка (токенизация и векторизация как популярные методы анализа текстовых данных; применение NLP для извлечения инсайтов из больших объемов текстовой информации)	2				3	ОПК-7.2, ПК-5.3
7.	Этика и безопасность в анализе больших данных при проведении внутреннего аудита системы управления качеством организации (проблемы конфиденциальности и защиты данных; этические аспекты работы с большими данными и алгоритмами)	2				3	ОПК-7.2, ПК-5.3
8.	Применение искусственного интеллекта в Big Data (AI и машинное обучение на смене подходов к анализу больших данных; кейсы успешного применения AI в различных отраслях)	2				2.85	ОПК-7.2, ПК-5.3
9.	Будущее Big Data: тренды и вызовы (новые технологии и подходы, такие как квантовые вычисления и edge computing; прогнозы развития Big Data и его влияние на бизнес и общество, Big Data при проведении внутреннего аудита системы управления качеством организации).	2		0.15	1	3	ОПК-7.2, ПК-5.3
Раздел 2. Анализ и визуализация больших данных. Язык R как один из стандартов при работе с данными и основа для повышения эффективности системы управления качеством организации							
1.	Введение в анализ данных и язык R (основные понятия и цели анализа данных, знакомство с языком R - установка и настройка среды разработки (RStudio), обзор основных		1			3	ОПК-7.2, ПК-5.3

	библиотек для анализа данных (dplyr, ggplot2, tidyr).						
2.	Основы работы с данными в R , в т.ч. при проведении внутреннего аудита системы управления качеством организации (загрузка и импорт данных из различных форматов, например, CSV, Excel, базы данных; основные структуры данных в R - векторы, списки, матрицы, датафреймы; основные операции с данными: фильтрация, сортировка, агрегация.		1			3	ОПК-7.2, ПК-5.3
3.	Предварительная обработка данных (очистка данных - работа с пропущенными значениями и дубликатами; преобразование данных - изменение типов, создание новых переменных; использование библиотеки dplyr для манипуляции данными)		1			3	ОПК-7.2, ПК-5.3
4.	Визуализация данных с помощью ggplot2 (основы визуализации - важность графиков в анализе данных; создание и настройка базовых графиков)		1			3	ОПК-7.2, ПК-5.3
5.	Расширенные методы визуализации при проведении внутреннего аудита системы управления качеством организации (создание сложных графиков: многослойные визуализации и использование faceting; визуализация временных рядов и пространственных данных; интерактивная визуализация с помощью библиотеки plotly)		1			3	ОПК-7.2, ПК-5.3
6.	Основы статистического анализа при проведении внутреннего аудита системы управления качеством организации (введение в статистику: описательная статистика и основные концепции; параметрические и непараметрические тесты: t-тест, ANOVA, тесты на		1			3	ОПК-7.2, ПК-5.3

	корреляцию; применение статистических методов в R)						
7.	Моделирование данных (введение в линейную регрессию: основные понятия и интерпретация коэффициентов; построение моделей в R; оценка качества модели)		1			3	ОПК-7.2, ПК-5.3
8.	Работа с большими данными при проведении внутреннего аудита системы управления качеством организации (понятие больших данных и их особенности; использование пакетов для работы с большими объемами данных (data.table, dplyr); основы параллельных вычислений в R)		1			2.85	ОПК-7.2, ПК-5.3
9.	Проектный подход к анализу данных (разработка проекта по анализу данных: выбор темы, сбор данных, анализ и визуализация; подготовка отчетов и презентаций результатов анализа; обзор лучших практик и инструментов для совместной работы над проектами)		1	0.15	1	3	ОПК-7.2, ПК-5.3
	Контроль	34					
	Итого	36	18	0.3	2	53.7	

5. Материально-техническое и учебно-методическое обеспечение дисциплины

5.1 Литература:

Основная литература

1. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. — Москва : Издательство Юрайт, 2025. — 97 с. — (Высшее образование). — ISBN 978-5-534-21173-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559502>

2. Толстобров, А. П. Управление данными : учебник для вузов / А. П. Толстобров. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 272 с. — (Высшее образование). — ISBN 978-5-534-14162-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/567678>

Дополнительная литература

1. Гордеев, С. И. Организация баз данных : учебник для вузов / С. И. Гордеев, В. Н. Волошина. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2025. — 691 с. — (Высшее образование). — ISBN 978-5-534-21115-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559377>

5.2. Перечень лицензионного программного обеспечения

1. Astra Linux Special Edition «Смоленск», «Орел»; РедОС ; ОС "Альт Рабочая станция" 10; ОС "Альт Образование" 10

2. МойОфис Стандартный 2, МойОфис Образование, Р7-Офис Профессиональный, МойОфис Стандартный 3, МойОфис Профессиональный 3

5.3 Современные профессиональные базы данных, к которым обеспечивается доступ обучающихся

1. Профессиональная база данных «Информационные системы Министерства экономического развития Российской Федерации в сети Интернет» (Портал «Официальная Россия» - <http://www.gov.ru/>)
2. Государственная система правовой информации «Официальный интернет-портал правовой информации» (<http://pravo.gov.ru/>)
3. Профессиональная база данных «Финансово-экономические показатели Российской Федерации» (Официальный сайт Министерства финансов РФ - <https://www.minfin.ru/ru/>)
4. Профессиональная база данных «Официальная статистика» (Официальный сайт Федеральной службы государственной статистики - <http://www.gks.ru/>)

5.4. Информационно-справочные системы, к которым обеспечивается доступ обучающихся

1. Справочно-правовая система «Консультант Плюс»
2. Справочно-правовая система «ГАРАНТ-Максимум»

5.5. Специальные помещения

Учебные аудитории для проведения занятий лекционного типа	Комплекты ученической мебели Мультимедийный проектор Доска Экран
Учебные аудитории для проведения практических занятий (занятий семинарского типа)	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Учебные аудитории для групповых и индивидуальных консультаций	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Учебные аудитории для текущего контроля и промежуточной аттестации	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Помещения для самостоятельной работы	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Помещения для хранения и профилактического обслуживания оборудования	Комплекты специализированной мебели для хранения оборудования

6. Фонд оценочных средств по дисциплине Анализ больших данных

6.1. Типовые контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций

ОПК-7 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

№ п/п	Задание	Ключ к заданию / Эталонный ответ	Критерии оценивания
1	На какие платформы можно установить язык R для решения задач профессиональной деятельности? а) Windows, MacOS б) MacOS, Linux в) Windows, MacOS, Linux	в) Windows, MacOS, Linux	Указан единственно верный вариант ответа
2	Одним из важнейших ресурсов администрирования языка R – лидера в современных информационных технологиях, является система а) CRAT б) CRAN (Comprehensive R Archive Network) в) R демократичен	б) CRAN	Указан единственно верный вариант ответа
3	Что такое вектор? а) это одномерный набор данных, это как книги на полке, у каждого значения вектора есть своя позиция в этой последовательности, и это позиция обозначается квадратными скобками б) числа в ряд в) 100 числовых значений	а) одномерный набор данных, это как книги на полке, у каждого значения вектора есть своя позиция в этой последовательности, и это позиция обозначается квадратными скобками	Указан единственно верный вариант ответа
4	Выберите правильный вариант задания последовательности с помощью функции seq() а) my_vector_5 <- seq(from - 1, to - 10, by - 1) б) my_vector_6<- [seq(from-1, to-10, by-0.5)] в) my_vector_7<- seq[from-1, to-10, length.out-20]	в) my_vector_7<- seq[from-1, to-10, length.out-20]	Указан единственно верный вариант ответа
5	Какая из NoSQL СУБД для решения задач профессиональной деятельности подходит для организации консистентного и распределённого хранилища? а) Hbase б) Cassandra в) Redis	б) Cassandra	Указан единственно верный вариант ответа

	г) p> BigTable</p>		
6.	Установить соответствие: 1. Apache Hadoop 2. Машинное обучение 3. ETL 4. Data Mining (добыча данных) 5. SQL 6. Нормализация базы данных 7. Big Data 8. Визуализация данных 9. Кластеризация 10. NoSQL А. Процесс извлечения, трансформации и загрузки данных В. Метод, используемый для группировки объектов по схожести С. Подход к обработке и анализу больших объемов данных Д. Язык для работы с реляционными базами данных Е. Инструмент для распределенного хранения и обработки данных Ф. Процесс уменьшения избыточности данных в базе Г. Анализ данных для выявления скрытых закономерностей Н. Способ представления данных в графическом формате И. Алгоритмы, позволяющие компьютерам учиться на данных Ж. Тип базы данных, не использующий традиционные таблицы	1 - Е 2 - И 3 - А 4 - G 5 - D 6 - F 7 - C 8 - Н 9 - В 10 - J	Указаны все верные варианты ответов

6	Установить соответствие. 1) между величинами X и Y существует положительная связь 2) между величинами X и Y существует отрицательная связь 3) между величинами X и Y не существует связи а) парный коэффициент корреляции ρ от -1 до 0 б) значение коэффициента корреляции близко к 0 в) парный коэффициент корреляции ρ от 0 до 1	1 – в, 2 – а, с - б	Указаны все верные варианты ответов
7	Установить соответствие. 1. Количественная характеристика продукции... 2. Количественная характеристика процесса... а) числовая оценка параметров; б) единичные показатели; в) простые показатели; г) суммарные показатели.	1 – б, 2 – а.	Указаны все верные варианты ответов
8	Установите правильную последовательность при выполнении основных этапов процесса анализа больших данных для решения задач профессиональной деятельности. Анализ данных → Сбор данных → Предобработка данных → Визуализация результатов → Интерпретация результатов	Сбор данных → Предобработка данных → Анализ данных → Визуализация результатов → Интерпретация результатов.	Дан верный ответ
9	Установите правильную последовательность в шагах в построении модели машинного обучения для анализа больших данных для решения задач профессиональной деятельности? Обучение модели → Сбор данных → Оценка модели → Предобработка данных → Внедрение модели	Сбор данных → Предобработка данных → Обучение модели → Оценка модели → Внедрение модели	Дан верный ответ
10	Для решения задач профессиональной деятельности ответьте на вопрос, как называется нарушение допущения о постоянстве дисперсии остатков?	гетероскедастичность	Указан единственно верный вариант ответа

11	Для решения задач профессиональной деятельности ответьте на вопрос, как называется нарушение допущения о независимости остатков?	автокорреляция	Указан единственно верный вариант ответа
12	Для решения задач профессиональной деятельности ответьте на вопрос, как называется метод изучения объекта не непосредственно, а через рассмотрение подобного ему и более простого объекта (т.е. способ изучения объекта путем создания и исследования его упрощенной копии)?	моделирования	Указан единственно верный вариант ответа
13	При работе с современными информационными технологиями к слабоструктурированным относятся задачи, которые содержат ... переменные	количественные, так и качественные	Указан единственно верный вариант ответа
14	При работе с современными информационными технологиям как Вы считаете, как получить справку по функции R?	Можно одним из двух действий: написать знак вопроса и название функции: ?function ; с помощью функции help: help(function)	Допустимые ответы: написать знак вопроса и название функции: ?function ; с помощью функции help
15	При работе с современными информационными технологиям как Вы считаете, как в R установить путь к рабочей директории "Учебный проект", созданной в папке "Мои документы" на диске C?	setwd("C:/Мои_документы/Учебный проект")	Указан единственно верный вариант ответа
16	При работе с современными информационными технологиями, постройте для набора данных quakes пакета datasets гистограмму распределения глубин	ggplot(quakes, aes(x = depth)) + geom_histogram(binwidth = 5, fill = "blue", color = "black") + labs(title = "Гистограмма распределения глубин", x = "Глубина (метры)", y = "Частота") + theme_minimal()	Указан единственно верный вариант ответа
17	При работе с современными информационными технологиями, постройте для набора данных quakes пакета datasets гистограмму распределения магнитуд	ggplot(quakes, aes(x = mag)) + geom_histogram(binwidth = 0.2, fill = "green", color = "black") + labs(title = "Гистограмма распределения магнитуд", x = "Магнитуда", y = "Частота") + theme_minimal()	Указан единственно верный вариант ответа

ПК-5 - Способен проводить внутренний аудит системы управления качеством организации, анализировать его результаты, разрабатывать мероприятия по предотвращению выпуска некачественной продукции

№ п/п	Задание	Ключ к заданию / Эталонный ответ	Критерии оценивания
1	При проведении внутреннего аудита системы управления качеством организации, открывая тот или иной файл, вы видите, что этот файл выглядит как набор непонятных символов, как это исправить? а) изначальный файл пересохранить как txt б) установить рабочую директорию в) изменить кодировку на UTF-8	в) изменить кодировку на UTF-8	Указан единственно верный вариант ответа
2	Что такое рабочая директория а) относительный путь б) папка в операционной системе в) абсолютный путь	б) папка в операционной системе	Указан единственно верный вариант ответа
3	Что такое Stack Owerflow а) сообщество форума б) справка по r в) сайт администрирования r	а) сообщество форума	Указан единственно верный вариант ответа
4	При проведении внутреннего аудита системы управления качеством организации, инвестиционный фонд интересуется тем, почему часть финансируемых им проектов успешно переходят на второй год, а часть — нет. К какому типу относится эта задача анализа данных? а) построение решающего правила б) классификация в) поиск информативных признаков г) цензурирование	б) классификация	Указан единственно верный вариант ответа
5	При проведении внутреннего аудита системы управления качеством организации, какая функция в R позволяет повторить вектор или его элемент: а) rep[,] б) rep(1:5, 4) в) Rep(1:5, 4).	б) rep(1:5, 4)	Указан единственно верный вариант ответа

6	При проведении внутреннего аудита системы управления качеством организации, установите соответствие между методами анализа данных и их описаний 1 Кластеризация 2. Регрессия 3. Ассоциативные правила 4. Классификация а. Метод, используемый для прогнозирования числовых значений на основе входных данных. б. Метод, который группирует объекты по схожести, не требуя заранее заданных меток. в. Метод, который определяет взаимосвязи между переменными и находит часто встречающиеся комбинации элементов. г. Метод, который присваивает метки объектам на основе обучающего набора данных.	1. - б. 2. - а. 3. - в. 4. - г.	Указаны все верные варианты ответов
7	При проведении внутреннего аудита системы управления качеством организации, установите соответствие. 1. Большие данные 2. Машинное обучение 3. Непрерывные данные 4. Структурированные данные а. Данные, которые имеют четкую организацию и легко поддаются анализу с помощью традиционных инструментов. б. Процесс, в ходе которого алгоритмы обучаются на исторических данных для предсказания будущих событий. в. Данные, которые могут принимать любое значение в пределах заданного диапазона. г. Объем данных, который превышает возможности традиционных методов обработки и анализа	1. - г. 2. - б. 3. - в. 4. - а.	Указаны все верные варианты ответов
8	При проведении внутреннего аудита системы управления качеством организации, установите правильную последовательность. Единицей измерения количества информации по возрастанию используются следующие: гигабайт (Гб), байт, мегабайт (Мбайт), килобайт (Кбайт)	байт, килобайт (Кбайт), мегабайт (Мбайт), гигабайт (Гб)	Дан верный ответ
9	При проведении внутреннего аудита системы управления качеством	5→ 6→ 3→ 2→ 10→ 4→ 7→ 1→ 8→ 9	Дан верный ответ

	организации, установите правильную последовательность. 1. Внедрение изменений. 2. Анализ данных. 3. Очистка и подготовка данных. 4. Выявление причин проблем. 5. Определение целей и задач аудита. 6. Сбор данных. 7. Разработка рекомендаций. 8. Разработка рекомендаций 9. Обратная связь и корректировка. 10. Оценка соответствия стандартам		
10	У каждого элемента вектора есть своя позиция в последовательности. Как эта позиция обозначается в среде R	с помощью квадратных скобок []	Указан единственно верный вариант ответа
11	К какому типу шкал относится шкала «очень плохо»-«плохо»-«средне»-«хорошо»-«очень хорошо»?	порядковая шкала	Указан единственно верный вариант ответа
12	При проведении внутреннего аудита системы управления качеством организации,, что можно сделать с помощью команды <code>my_file.read()</code>	прочитать весь файл	Указан единственно верный вариант ответа
13	При проведении внутреннего аудита системы управления качеством организации, для чего используется команда <code>readline()</code>?	для чтения одной строки из файла	Указан единственно верный вариант ответа
14	Имеется множество <code>a</code> - {1, 3, 5, 7}. При проведении внутреннего аудита системы управления качеством организации, какой оно примет вид после выполнения команды <code>a.remove(7)</code>	{1, 3, 5}	Указан единственно верный вариант ответа
15	Имеется множество <code>a</code> - {1, 3, 5, 7}. При проведении внутреннего аудита системы управления качеством организации, какой оно примет вид после выполнения команды <code>a.add(2)</code>	{1, 2, 3, 5, 7}	Указан единственно верный вариант ответа
16	Имеются примеры значений <code>values <- c(1, -3.5, 0, Inf, -Inf, NaN)</code>. При проведении внутреннего аудита системы управления качеством организации, что вернет функция <code>is.finite</code> в R	TRUE TRUE TRUE FALSE FALSE FALSE	Указан единственно верный вариант ответа
17	При проведении внутреннего аудита системы управления качеством организации, в программе выполните импорт модуля <code>math</code> под именем <code>m</code>. Далее пропишите пример корректного обращения к функции <code>sin</code>	<code>import math as m</code> <code>angle_degrees = 30</code> <code>result = m.sin(angle_degrees)</code>	Допустимые ответы : при значении <code>angle_degrees</code> (угол в градусах) от 0 до 180

6.2 Комплект оценочных средств для промежуточной аттестации

Примерные вопросы к экзамену

ПК-5 - Способен проводить внутренний аудит системы управления качеством организации, анализировать его результаты, разрабатывать мероприятия по предотвращению выпуска некачественной продукции

№п/п	Вопрос	Ключ к заданию / Эталонный ответ
1	Основные характеристики больших данных	Основные характеристики Big Data: объем (огромные объемы информации вплоть до терабайтов и петабайтов), скорость (высочайшая, при которой возможно обрабатывать информацию в режиме реального времени), разнообразие (текстовые данные, изображения, видео, аудио и т.д.), достоверность (высокое качество данных), ценность (для бизнеса и науки), сложность
2	Применение Big Data в различных отраслях	Финансовый сектор (риск-менеджмент и алгоритмическая торговля), здравоохранение (персонализированная медицина, мониторинг здоровья пациентов), розничная торговля (анализ поведения клиентов, управление запасами), производственный сектор (предиктивное обслуживание, оптимизация производственных процессов), транспорт и логистика (управление цепочками поставок, анализ трафика), телекоммуникации (улучшение качества обслуживания, управление сетями), гос. сектор (анализ общественного мнения: управление городским хозяйством), спорт и развлечения (анализ производительности спортсменов, персонализированные рекомендации).
3	Технологии хранения данных Hadoop	Hadoop стал стандартом в области обработки Big Data благодаря своей способности эффективно хранить и обрабатывать огромные объемы информации. Его компоненты (HDFS, YARN, MapReduce, Apache Hive, Apache HBase, Apache Parquet и ORC) позволяют легко разрабатывать и запускать приложения для анализа данных в распределенной среде, что делает его популярным выбором среди организаций, работающих с большими данными.
4	NoSQL базы данных	NoSQL базы данных предоставляют гибкие и масштабируемые решения для хранения и обработки данных в условиях больших объемов и разнообразия информации. Отлично подходят для современных приложений, требующих высокой производительности и адаптивности. Выбор конкретной NoSQL базы данных зависит от типа данных, требований проекта и необходимых функций.
5	Сравнение традиционных реляционных баз данных и NoSQL	Выбор между реляционными и NoSQL базами данных зависит от специфики проекта, требований к данным, масштабируемости и производительности. Реляционные базы данных обеспечивают высокую степень согласованности и подходят для структурированных данных, в то время как NoSQL базы данных предлагают гибкость и масштабируемость для работы с неструктурированными и полуструктурированными данным
6	ML - машинное обучение (supervised и unsupervised методы)	Это область искусственного интеллекта, занимающаяся разработкой алгоритмов и моделей, позволяющих компьютерам обучаться на основе данных и делать предсказания без явного программирования. В ML выделяют два основных подхода: обучение с учителем (модель обучается на размеченных данных) и обучение без учителя (у модели нет меток, и она должна самостоятельно выявлять структуры и закономерности в данных).
7	Деревья решений в ML	Популярный метод ML, используется как для задач классификации, так и для регрессии. Характеристики: 1. Структура имеет древовидную структуру, состоящую из узлов (содержит весь набор данных), ветвей (признаки данных) и

		листьев (конечные решения или классы). 2. Разделение данных. 3. Интерпретируемость. Известные алгоритмы для построения деревьев решений включают ID3, C4.5, CART.
8	Токенизация и векторизация	Являются важными шагами при анализе текстовых данных, помогают преобразовать неструктурированные тексты в структурированные форматы, которые могут быть использованы для построения моделей машинного обучения. Правильный выбор методов токенизации и векторизации может существенно повлиять на качество анализа и результаты моделирования.
9	Применение NLP для извлечения инсайтов из больших объемов текстовой информации	Открывает новые возможности для анализа данных и принятия решений. Эти методы помогают организациям лучше понимать своих клиентов, оптимизировать бизнес-процессы и выявлять новые возможности. Некоторые подходы и методы, используемые для этой цели: анализ тональности, извлечение ключевых слов и тем, классификация текста, извлечение информации, сущностное распознавание, суммаризация текста, анализ социальных сетей, чат-боты и виртуальные помощники.
10	Принципы эффективной визуализации больших данных	1. Определение цели. 2. Простота и ясность. 3. Подбор правильных типов визуализаций. 4. Контекст и интерпретация. 5. Использование цвета. 6. Интерактивность. 7. Адаптивность. 8. Тестирование и обратная связь. 9. Обновление данных.
11	Инструмент для визуализации данных Tableau	Один из самых популярных инструментов для визуализации данных, позволяющий создавать интерактивные дашборды и отчеты. Особенности: поддержка различных источников данных, возможность работы с большими объемами информации, простота в использовании.
12	Инструмент для визуализации данных Power BI	Этот инструмент от Microsoft для анализа и визуализации данных, поэтому главная особенность - интеграция с другими продуктами Microsoft, возможность создания интерактивных отчетов и дашбордов.
13	Инструмент для визуализации данных D3.js	JavaScript-библиотека для создания динамических и интерактивных визуализаций в веб-приложениях. Особенности являются высокая гибкость и возможность создания уникальных графиков.
14	Инструмент для визуализации данных Google Data Studio	Бесплатный инструмент от Google для создания отчетов и дашбордов. поэтому главная особенность - простейшая интеграция с другими сервисами Google и простота в использовании.
15	Основные концепции, используемые потоковой обработки Apache Kafka	1) Сохранение данных: Kafka сохраняет все сообщения на диске, что позволяет восстанавливать данные в случае сбоя. 2) Гарантии доставки: Kafka предоставляет различные уровни гарантии доставки сообщений такие как At-most-once, At-least-once, Exactly-once:
16	Архитектура Apache Kafka	1) Kafka работает в виде кластера, состоящего из нескольких брокеров, что обеспечивает высокую доступность и отказоустойчивость. 2) ZooKeeper: Используется для управления метаданными и координации между брокерами.

ОПК-7 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

№п/п	Вопрос	Эталонный ответ
------	--------	-----------------

1	Какие типы данных поддерживаются в R? Каковы их англоязычные наименования?	В R поддерживаются следующие типы данных: numeric (числовые), integer (целые числа), character (строки), logical (логические), complex (комплексные числа), list (списки), data.frame (таблицы).
2	Что такое переменная?	Это именованная область памяти, в которой хранится значение. Она используется для хранения данных и может быть изменена в процессе выполнения программы.
3	Какой оператор используется для записи значения в переменную?	Оператор присваивания = или <-. Например: x <- 10 или x = 10.
4	С помощью какого семейства функций можно преобразовывать типы переменных?	Функции преобразования типов: as.numeric(), as.integer(), as.character(), as.logical(), as.complex().
5	Можно ли использовать ранее созданное имя переменной для хранения новых данных другого типа?	В R можно переопределять переменные, присваивая им значения другого типа.
6	Можно ли записать в переменную результат выполнения выражения, в котором она сама же и участвует?	Возможно. Например: x <- 5; x <- x + 2 изменит значение x на 7.
7	Какая функция позволяет прочесть пользовательский ввод с клавиатуры в консоли? Какой тип данных будет иметь возвращаемое значение?	Функция readline(). Возвращаемое значение имеет тип character.
8	Какую функцию можно использовать для вывода значения переменной в консоль? Чем отличается использование этой функции от случая, когда вы просто пишете название переменной в строке программы?	Функция print(). Использование print() явно выводит значение, тогда как просто написание имени переменной может не вывести значение в некоторых случаях (например, в функциях).
9	Какой символ является разделителем целой и дробной части при записи чисел с плавающей точкой?	В R используется точка (.) как разделитель целой и дробной части.
10	Что такое операторы и операнды? Приведите примеры бинарных и унарных операторов.	Операторы — это символы, которые выполняют операции над операндами. Бинарные операторы (например, +, -, *, /) работают с двумя операндами, унарные (например, -, !) — с одним.
11	Какое значение будет иметь результат деления на ноль?	Результат деления на ноль в R будет Inf (бесконечность) или NaN (не число) в зависимости от контекста.
12	Какие функции выполняют операторы %%, %/%, ^, **?	%*% — матричное умножение, %/% — целочисленное деление, ^ и ** — возведение в степень.
13	Как проверить, является ли число четным?	Для проверки четности числа можно использовать выражение: x %% 2 == 0, где x — проверяемое число.
14	Как определить количество символов в строке?	Для определения количества символов в строке используется функция nchar().

15	Как называется операция состыковки нескольких строк и с помощью какой функции она выполняется? Как добиться того, чтобы при этом не добавлялись пробелы между строками?	Операция называется конкатенация. Она выполняется с помощью функции <code>paste()</code> . Чтобы избежать пробелов между строками, нужно использовать аргумент <code>sep = ""</code> : <code>paste("Hello", "World", sep = "")</code> .
----	---	---

6.3 Методические материалы, определяющие критерии оценивания сформированности компетенций

Критерии и шкалы оценивания промежуточной аттестации Шкала и критерии оценки (экзамен)

Отлично	Хорошо	Удовлетворительно	Неудовлетворительно
1. Полно раскрыто содержание вопросов билета. 2. Материал изложен грамотно, в определенной логической последовательности, правильно используется терминология. 3. Показано умение иллюстрировать теоретические положения конкретными примерами, применять их в новой ситуации. 4. Продemonстрировано усвоение ранее изученных сопутствующих вопросов, сформированность умений и знаний. 5. Ответ прозвучал самостоятельно, без наводящих вопросов.	1. Ответ удовлетворяет в основном требованиям на оценку «5», но при этом может иметь следующие недостатки: в изложении допущены небольшие пробелы, не исказившие содержание ответа. 2. Допущены один - два недочета при освещении основного содержания ответа, исправленные по замечанию экзаменатора. 3. Допущены ошибка или более двух недочетов при освещении второстепенных вопросов, которые легко исправляются по замечанию экзаменатора.	1. Неполно или непоследовательно раскрыто содержание материала, но показано общее понимание вопроса и продемонстрированы умения, достаточные для дальнейшего усвоения материала. 2. Имелись затруднения или допущены ошибки в определении понятий, использовании терминологии, исправленные после нескольких наводящих вопросов. 3. При неполном знании теоретического материала выявлена недостаточная сформированность умений и знаний.	1. Содержание материала нераскрыто. 2. Ошибки в определении понятий, не использовалась терминология в ответе.