Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гриб Владислав Валерьевич
Должность: Ректор
Дата подписания: 14.10.2026 17:08:07
Уникальный программный ключ:
637517d24e103c3db032acf37e839d98ec1c5bb2f5eb89c29abfcd7f4398544

Образовательное частное учреждение высшего образования
«МОСКОВСКИЙ УНИВЕРСИТЕТ ИМЕНИ А.С. ГРИБОЕДОВА»
(ИМПЭ им. А.С. Грибоедова)

МЕЖДУНАРОДНЫЙ ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И БИЗНЕС-ИНФОРМАТИКИ

УТВЕРЖДАЮ
Директор Международного института информационных технологий и бизнес-информатики
_____ Т.В. Новикова
«17» декабря 2025г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по учебной дисциплине:
ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ СТАТИСТИЧЕСКОГО АНАЛИЗА

направление подготовки
38.03.05
БИЗНЕС-ИНФОРМАТИКА
(уровень бакалавриат)

направленность (профиль):
«Цифровая трансформация бизнеса»

Форма обучения: очная, очно-заочная

Москва

Фонд оценочных средств для дисциплины «Программное обеспечение статистического анализа». Направление подготовки/специальность 09.03.03 Прикладная информатика, направленность (профиль/специализация): Анализ Данных / О.А. Левичев – М.: ИМПЭ им. А.С. Грибоедова.

Фонд оценочных средств является неотъемлемой частью рабочей программы дисциплины.

Разработчик: Т.В. Новикова

Заведующий кафедрой _____/Т.В. Новикова, к. э. н., доцент

1. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

Настоящий Фонд оценочных средств (ФОС) по дисциплине «Программное обеспечение статистического анализа» является неотъемлемым приложением к рабочей программе дисциплины (РПД) «Программное обеспечение статистического анализа». На данный ФОС распространяются все реквизиты утверждения, представленные в РПД по данной дисциплине.

2. ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ

Для определения качества освоения обучающимися учебного материала по дисциплине используются следующие оценочные средства:

№ п/п	Оценочное средство	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
1	Тестирование	Вид контроля, позволяющий оценить изученный теоретический материал.	Вопросы для проведения тестирования
2	Практические задания	Вид контроля, позволяющий оценить умение обучающегося применять осваиваемую компетенцию в практических ситуациях и при решении производственных задач	Задания к практическому занятию
3	Контрольная работа	Вид контроля, позволяющий определить результат освоения компетенций по дисциплине в рамках рассматриваемой темы, оцениваемый с помощью соответствующих индикаторов достижения компетенций	Задания контрольной работы
4	Самостоятельная работа	Вид контроля, позволяющий оценить проработку теоретического материала, изучение рекомендуемой литературы, выполнение практико-ориентированных заданий (заполнение таблиц, проведение сравнительного анализа, составление схем и др.), решение практических задач, создание презентаций, написание рефератов, подборку нормативного и иного материала и выполнение других заданий	Задания самостоятельной работы
5	Курсовая работа	Вид контроля, позволяющий выявить степень владения базовыми знаниями, умениями и навыками, необходимыми для обучения, и определить уровень владения новым материалом	Индивидуальные задания (темы) для курсовой работы
6	Зачет/Зачет с оценкой/Экзамен	Вид контроля, позволяющий выявить степень овладения знаниями, умениями и навыками, необходимыми для дальнейшего освоения образовательной программы подготовки	Вопросы для подготовки к зачету/зачету с оценкой/экзамену

3. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ ДИСЦИПЛИНЕ

3.1. Сопроводительная информация.

Разработчик	Т.В. Новикова
Кафедра	Бизнес-информатика технологий управления
Наименование дисциплины	Программное обеспечение статистического анализа
Факультет / институт	Международный институт информационных технологий и бизнес-информатики
Направление подготовки / специальность	38.03.05 Бизнес-информатика
Количество вопросов в оценочных заданиях (диапазон)	От 15 до 30
Общее время тестирования (мин)	90
Общее количество вопросов/заданий в ФОС	30
Размещенность на сайте Университета примерного перечня вопросов, заданий ФОС – для подготовки обучающихся к прохождению оценки (да / нет)	Да

3.2. Характеристика оцениваемых компетенций

Код компетенции	Формулировка компетенции	Индикаторы достижения компетенции
ПК-1	Способен осуществлять сбор и обработку исходных данных для составления проектов финансово-хозяйственной, производственной и коммерческой деятельности (бизнес-планов) организации	**ИПК-1.1 Знать**: Нормативные правовые акты, регулирующие финансово-хозяйственную деятельность организации Методы сбора и обработки экономической информации, а также осуществления технико-экономических расчетов и анализа хозяйственной деятельности организации, с использованием вычислительной техники Порядок разработки бизнес-планов организации в соответствии с отраслевой направленностью **ИПК-1.2 Уметь**: Использовать автоматизированные системы сбора и обработки экономической информации Применять информационные технологии для обработки экономических данных Собирать и анализировать исходные данные, необходимые для расчета экономических и финансово-экономических показателей, характеризующих деятельность организации **ИПК-1.3 Владеть**: Навыками сбора и обработки исходных данных для составления проектов финансово-хозяйственной, производственной и коммерческой деятельности (бизнес-планов) организации

4. СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ТЕКУЩЕГО КОНТРОЛЯ

4.1. ТИПОВЫЕ ТЕСТОВЫЕ ЗАДАНИЯ.

Тесты содержат набор вопросов, в полном объеме охватывающие изученный теоретический материал по указанной теме. Выполнение тестов позволяет определить результат освоения компетенций по дисциплине в рамках рассматриваемой темы, оцениваемый с помощью соответствующих индикаторов достижения компетенций. Индивидуальный тестовый сеанс для каждого обучающегося формируется по специальному алгоритму, обеспечивающему заданную тематическую структуру и пропорциональное наличие вопросов разного типа и сложности.

При формировании тестов необходимо использовать задания следующих типов:

Тип задания 1. Задания закрытого типа на установление соответствия.

Тип задания 2. Задания закрытого типа на установление последовательности.

Тип задания 3. Задания комбинированного типа, предполагающие выбор одного правильного ответа из предложенных

Тип задания 4. Задания комбинированного типа, предполагающие выбор нескольких ответов из предложенных

Тип задания 5. Задания открытого типа с развернутым ответом.

<table>
<tr><th>Тема</th><th>Код и формулировка компетенции</th><th>Индикаторы достижения компетенции</th><th>Тест по теме</th></tr>
<tr><td>Тема 1. Визуализация данных с использованием пакетов статистического анализа.</td><td>ПК-1</td><td>ИПК-1.1</td><td>
Задания закрытого типа на установление соответствия.
Знание темы визуализация данных с использованием пакетов статистического анализа необходимо для эффективного представления сложных количественных результатов в наглядной и интерпретируемой форме, что особенно важно при принятии обоснованных управленческих и аналитических решений. Установите соответствие между типом визуализации и её наиболее подходящим применением в рамках методологии проектирования аналитических дашбордов.
<table>
<tr><td></td><td>Тип визуализации</td><td></td><td>Применение</td></tr>
<tr><td>А</td><td>Гистограмма</td><td>1</td><td>Отображение распределения категориальных данных</td></tr>
<tr><td>Б</td><td>Круговая диаграмма</td><td>2</td><td>Показ соотношения частей к целому при небольшом числе категорий</td></tr>
<tr><td>В</td><td>Диаграмма рассеяния (scatter plot)</td><td>3</td><td>Анализ формы распределения количественной переменной</td></tr>
<tr><td>Г</td><td>Столбчатая диаграмма</td><td>4</td><td>Выявление корреляции между двумя числовыми переменными</td></tr>
</table>
Ответ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td>3</td><td>2</td><td>4</td><td>1</td></tr>
</table>
Задания закрытого типа на установление соответствия.
Знание методов и средств проектирования баз данных в сочетании с навыками визуализации с использованием пакетов статистического анализа необходимо для обеспечения целостности, эффективного извлечения и наглядного
</td></tr>
</table>

<table>
<tr><th>Тема</th><th>Код и формулировка компетенции</th><th>Индикаторы достижения компетенции</th><th>Тест по теме</th></tr>
<tr><td></td><td></td><td></td><td>представления данных на всех этапах аналитического цикла — от хранения до принятия решений.
Установите соответствие между пакетом статистического анализа и функцией, используемой для создания графиков в соответствии с принципами визуального проектирования.
<table>
<tr><td></td><td>Пакет / библиотека</td><td></td><td>Функция / инструмент визуализации</td></tr>
<tr><td>А</td><td>R (ggplot2)</td><td>1</td><td>sns.heatmap()</td></tr>
<tr><td>Б</td><td>Python (Matplotlib)</td><td>2</td><td>geom_bar()</td></tr>
<tr><td>В</td><td>Python (Seaborn)</td><td>3</td><td>plt.scatter()</td></tr>
<tr><td>Г</td><td>R (base graphics)</td><td>4</td><td>barplot()</td></tr>
</table>
Ответ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td>2</td><td>3</td><td>1</td><td>4</td></tr>
</table>

Задания закрытого типа на установление последовательности.
Установите правильную последовательность этапов работы с данными — от проектирования базы данных до визуализации результатов анализа с использованием статистических пакетов
1. Определение сущностей и связей, создание ER-диаграммы
2. Нормализация таблиц и реализация реляционной структуры в СУБД
3. Извлечение и агрегация данных с помощью SQL-запросов
4. Загрузка данных в среду анализа (например, R или Python)
5. Построение графиков (например, с помощью ggplot2 или seaborn) для интерпретации результатов
Ответ:
<table>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Тема</th><th>Код и формулировка компетенции</th><th>Индикаторы достижения компетенции</th><th>Тест по теме</th></tr>
<tr><td></td><td></td><td></td><td>
Задания закрытого типа на установление последовательности.
Расположите в логической последовательности действия, которые выполняет аналитик при подготовке отчёта на основе данных из спроектированной базы, используя пакеты статистического анализа.

Проектирование логической модели базы данных с учётом будущих аналитических запросов
Реализация физической модели базы данных и наполнение её данными
Экспорт выборки данных в формат, совместимый со статистическим пакетом (CSV, Parquet и др.)
Предварительная обработка и трансформация данных в R/Python
Визуализация ключевых метрик с применением принципов эффективного графического дизайна

Ответ:
<table>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
</table>
Задания комбинированного типа, предполагающие выбор одного правильного ответа из предложенных
При разработке системы аналитики продаж компания планирует визуализировать динамику выручки по регионам и категориям товаров за последние три года. Какой подход к проектированию базы данных наиболее рационален для обеспечения эффективной последующей визуализации в R или Python?
A) Хранение всех данных в одной неструктурированной таблице без нормализации
B) Использование денормализованной «плоской» таблицы, специально подготовленной для аналитических запросов (например, в формате звездной
</td></tr>
</table>

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			схемы) C) Полная нормализация до пятой нормальной формы с множеством связанных таблиц D) Отказ от реляционной модели в пользу хранения данных исключительно в JSON-файлах Ответ: B Обоснование: Звездная схема (денормализованная структура с фактами и измерениями) оптимизирует выполнение агрегирующих запросов, упрощает извлечение данных и минимизирует сложность JOIN-операций, что позволяет быстрее загружать согласованные наборы данных в среды визуализации (например, pandas + seaborn или ggplot2), обеспечивая корректность и производительность аналитических графиков. ***Задания комбинированного типа, предполагающие выбор одного правильного ответа из предложенных*** Аналитик получил задачу визуализировать частоту обращений клиентов по типам проблем, с разбивкой по месяцам и каналам поддержки. На каком этапе проектирования базы данных закладываются предпосылки для корректной и гибкой визуализации таких данных? A) На этапе выбора СУБД (например, MySQL vs PostgreSQL) B) На этапе логического проектирования — при определении сущностей, атрибутов и связей C) На этапе физического размещения базы данных на сервере D) На этапе написания скриптов визуализации в R или Python Ответ: B Обоснование: Именно на этапе логического проектирования определяется структура данных: выделяются сущности (например, «Обращение», «Клиент», «Канал поддержки»), их атрибуты (включая дату, тип проблемы) и связи. Грамотно спроектированная модель обеспечивает полноту, непротиворечивость и удобство извлечения данных, что напрямую влияет на возможность

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			построения точных, детализированных и интерпретируемых визуализаций с помощью статистических пакетов. ***Задания комбинированного типа, предполагающие выбор нескольких ответов из предложенных*** Какие из перечисленных характеристик проектирования базы данных напрямую влияют на качество и эффективность последующей визуализации данных с использованием пакетов статистического анализа (например, ggplot2, seaborn, matplotlib)? A) Наличие четко определённых внешних ключей между таблицами B) Использование осмысленных и согласованных названий столбцов C) Применение сжатия данных на уровне СУБД D) Нормализация или денормализация структуры в зависимости от целей анализа E) Хранение всех данных в формате BLOB Правильные ответы: A, B, D Обоснование: A: Четкие внешние ключи обеспечивают корректные связи между сущностями (например, «Клиент» — «Заказ»), что позволяет точно агрегировать и объединять данные перед визуализацией. B: Согласованные и понятные названия столбцов упрощают загрузку данных в аналитические среды (R/Python) и снижают риск ошибок при построении графиков. D: Выбор между нормализацией (для целостности) и денормализацией (для скорости аналитических запросов, например, в звёздной схеме) напрямую влияет на производительность и удобство извлечения данных, используемых в визуализации. ***Задания комбинированного типа, предполагающие выбор нескольких ответов из предложенных***

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			При подготовке данных из реляционной базы для визуализации временных трендов (например, ежедневных продаж) какие из следующих действий на этапе проектирования или подготовки данных являются целесообразными? A) Создание отдельной таблицы-измерения «Дата» с атрибутами (день недели, месяц, квартал, праздник/нет) B) Хранение даты как текстовой строки в произвольном формате (например, «22 января 2025») C) Обеспечение полноты временного ряда (включая дни без продаж) через LEFT JOIN с календарной таблицей D) Использование триггеров для автоматической визуализации новых записей E) Применение типов данных DATE/DATETIME для временных меток Ответ: A, C, E Обоснование: A: Таблица-измерение «Дата» (как часть измерительной модели) упрощает группировку и фильтрацию по временным признакам, что критично для построения осмысленных временных графиков. C: Заполнение пропусков в временных рядах (например, дней без продаж) предотвращает искажение визуализации — линейные графики будут отображать реальную динамику, а не «скачки» из-за отсутствующих точек. E: Использование стандартных временных типов (DATE/DATETIME) обеспечивает корректную сортировку, агрегацию и совместимость с функциями визуализации в R/Python (например, pd.to_datetime()). ***Задания открытого типа с развернутым ответом.*** Почему при проектировании базы данных важно учитывать будущие задачи визуализации, и как ошибки на этапе проектирования могут повлиять на качество графиков, создаваемых в R или Python? Ответ: При проектировании базы данных важно заранее учитывать аналитические и визуализационные цели, поскольку от структуры хранения зависит полнота, согласованность и доступность данных. Например, отсутствие

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			нормализованной таблицы «Время» или некорректное хранение даты как текста затрудняет построение временных рядов. Аналогично, дублирование значений, отсутствие внешних ключей или несогласованные категории (например, «Москва» и «г. Москва») приводят к искажению агрегированных показателей и, как следствие, к ошибочным выводам на графиках. Таким образом, грамотное проектирование — основа достоверной и эффективной визуализации. ***Задания открытого типа с развернутым ответом.*** Как использование измерительных моделей (например, звёздной схемы) при проектировании базы данных упрощает процесс визуализации данных с помощью пакетов статистического анализа, таких как ggplot2 или seaborn? Ответ: Звёздная схема, состоящая из центральной таблицы фактов и окружающих её таблиц измерений (dimensions), оптимизирует выполнение аналитических запросов: она минимизирует сложность JOIN-операций и обеспечивает быстрое извлечение агрегированных данных с четкой семантикой (например, продажи по регионам, категориям, периодам). Это позволяет легко загружать структурированные, готовые к анализу датафреймы в R или Python, где пакеты вроде ggplot2 или seaborn требуют именно «плоских» табличных данных с ясными переменными. Благодаря этому аналитик тратит меньше времени на очистку и преобразование, а визуализации получаются точными, гибкими и интерпретируемыми.
Тема 2. Решение задач статистического анализа с использованием программных пакетов.	**ПК-1**	**ИПК-1.1**	***Задания закрытого типа на установление соответствия.*** Знание этапов проектирования базы данных необходимо, поскольку оно обеспечивает структурную целостность, согласованность и доступность данных, что напрямую определяет качество, точность и эффективность решения задач статистического анализа с использованием программных пакетов, таких как R и Python. Установите соответствие между этапом проектирования базы данных и его вкладом в успешное решение задач статистического анализа с использованием программных пакетов (R, Python и

<table>
<tr><th>Тема</th><th>Код и формулировка компетенции</th><th>Индикаторы достижения компетенции</th><th>Тест по теме</th></tr>
<tr><td></td><td></td><td></td><td>
др.).
<table>
<tr><td></td><td>Этап проектирования БД</td><td></td><td>Вклад в статистический анализ</td></tr>
<tr><td>А</td><td>Концептуальное проектирование (ER-модель)</td><td>1</td><td>Обеспечивает целостность и однозначную интерпретацию переменных, используемых в моделях</td></tr>
<tr><td>Б</td><td>Логическое проектирование (реляционная схема)</td><td>2</td><td>Позволяет точно определить сущности и связи, необходимые для формирования аналитических выборок</td></tr>
<tr><td>В</td><td>Физическое проектирование (индексы, типы данных)</td><td>3</td><td>Ускоряет извлечение больших объёмов данных, что критично при загрузке в pandas или R</td></tr>
<tr><td>Г</td><td>Нормализация таблиц</td><td>4</td><td>Устраняет избыточность и аномалии, повышая качество данных для последующего анализа</td></tr>
</table>
Ответ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td>2</td><td>1</td><td>3</td><td>4</td></tr>
</table>
Знание особенностей проектирования базы данных необходимо, поскольку оно позволяет формировать структурированные, непротиворечивые и эффективно извлекаемые наборы данных, что напрямую повышает точность, производительность и интерпретируемость результатов статистического анализа, выполняемого с помощью программных пакетов, таких как R и Python. Установите соответствие между особенностью проектирования базы данных и преимуществом, которое она даёт при решении задач статистического анализа с
</td></tr>
</table>

<table>
<tr><th>Тема</th><th>Код и формулировка компетенции</th><th>Индикаторы достижения компетенции</th><th>Тест по теме</th></tr>
<tr><td></td><td></td><td></td><td>помощью программных пакетов.
<table>
<tr><td></td><td>Особенность проектирования БД</td><td></td><td>Преимущество для статистического анализа</td></tr>
<tr><td>А</td><td>Использование типов данных DATE/DATETIME</td><td>1</td><td>Позволяет корректно строить временные ряды и применять функции временной агрегации</td></tr>
<tr><td>Б</td><td>Создание внешних ключей между таблицами</td><td>2</td><td>Обеспечивает возможность безопасного объединения (JOIN) данных из разных сущностей</td></tr>
<tr><td>В</td><td>Применение доменных ограничений (CHECK, ENUM)</td><td>3</td><td>Снижает количество пропущенных или аномальных значений, упрощая предобработку</td></tr>
<tr><td>Г</td><td>Денормализация для аналитических нужд</td><td>4</td><td>Упрощает экспорт «плоской» таблицы, готовой к загрузке в pandas или data.frame</td></tr>
</table>
Ответ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td>1</td><td>2</td><td>в</td><td>4</td></tr>
</table>
Задания закрытого типа на установление последовательности.
Установите правильную последовательность этапов работы с данными — от проектирования базы данных до выполнения статистического анализа с использованием программных пакетов (например, R или Python).
1. Определение предметной области и выявление ключевых сущностей и связей (построение ER-модели)
2. Разработка реляционной схемы: нормализация, определение первичных и внешних ключей
3. Реализация физической модели БД в СУБД и наполнение данными
4. Извлечение релевантной выборки с помощью SQL-запросов</td></tr>
</table>

<table>
<tr><th>Тема</th><th>Код и формулировка компетенции</th><th>Индикаторы достижения компетенции</th><th>Тест по теме</th></tr>
<tr><td></td><td></td><td></td><td>5. Загрузка данных в среду анализа (например, в pandas DataFrame или R data.frame) и выполнение статистических расчётов
Ответ:<table><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table>

Задания закрытого типа на установление последовательности.
Расположите в логическом порядке действия, обеспечивающие корректное решение задачи регрессионного анализа на основе данных из корпоративной базы.
1. Проектирование таблицы фактов и измерений с учётом будущих аналитических признаков (например, «доход», «регион», «дата»)
2. Обеспечение целостности данных через ограничения (NOT NULL, FOREIGN KEY, CHECK)
3. Экспорт подготовленного набора данных в формат CSV или прямое подключение к БД из Python/R
4. Предварительная обработка данных (обработка пропусков, кодирование категорий) в среде статистического анализа
5. Построение и интерпретация регрессионной модели с использованием пакетов (statsmodels, scikit-learn, lm() в R)
Ответ:<table><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table>

Задания комбинированного типа, предполагающие выбор одного правильного ответа из предложенных
Аналитик планирует провести кластерный анализ клиентов на основе их покупательского поведения. Какой подход к проектированию базы данных наиболее рационален для обеспечения корректности и эффективности этого анализа при последующей работе в Python (с использованием pandas и scikit-learn)?</td></tr>
</table>

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			A) Хранение всех событий в одной неструктурированной таблице без внешних ключей B) Использование полностью нормализованной модели до пятой нормальной формы C) Применение денормализованной структуры (например, звёздной схемы) с центральной таблицей фактов и измерениями «Клиент», «Товар», «Время» D) Отказ от реляционной модели в пользу хранения данных только в JSON-файлах Выберите один правильный ответ. Ответ: C Обоснование: Звёздная схема обеспечивает баланс между целостностью данных и производительностью аналитических запросов. Таблица фактов содержит количественные метрики (например, сумма покупки, частота), а измерения — качественные признаки (регион, возраст, категория товара). Это позволяет легко агрегировать данные по клиентам и формировать готовый набор признаков для кластеризации в pandas. В отличие от полной нормализации (B), которая требует множества JOIN-запросов, или неструктурированного хранения (A, D), звёздная схема упрощает экспорт «плоской» таблицы, необходимой для алгоритмов машинного обучения. ***Задания комбинированного типа, предполагающие выбор одного правильного ответа из предложенных*** При решении задачи линейной регрессии в R (lm()) аналитику необходимо использовать данные о продажах, клиентах и рекламных кампаниях. На каком этапе проектирования базы данных закладываются основы для корректного формирования регрессионной выборки? A) На этапе выбора аппаратного обеспечения сервера СУБД B) На этапе логического проектирования — при определении сущностей, атрибутов и связей C) На этапе написания скрипта визуализации в ggplot2 D) На этапе настройки резервного копирования базы данных

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			Выберите один правильный ответ. Ответ: B Обоснование: Именно на этапе логического проектирования определяется, какие переменные (атрибуты) будут доступны для анализа, как они связаны между собой и как будут объединяться (через внешние ключи). Например, если связь между «продажей» и «рекламной кампанией» не предусмотрена, соответствующий предиктор окажется недоступен, что исказит модель. Грамотная логическая модель гарантирует полноту, непротиворечивость и однозначную интерпретацию признаков — ключевых условий построения валидной регрессионной модели. ***Задания комбинированного типа, предполагающие выбор нескольких ответов из предложенных*** Какие из перечисленных элементов проектирования базы данных напрямую способствуют повышению качества и эффективности решения задач статистического анализа (например, в R с помощью dplyr/lm() или в Python с помощью pandas/statsmodels)? A) Использование первичных и внешних ключей B) Применение доменных ограничений (CHECK, ENUM, NOT NULL) C) Хранение всех данных в одном CSV-файле без использования СУБД D) Наличие согласованных справочников (например, единый список регионов) E) Отказ от нормализации в пользу полной денормализации Ответ: A, B, D Обоснование: A: Первичные и внешние ключи обеспечивают ссылочную целостность, что критично при объединении таблиц (JOIN) для формирования полной аналитической выборки. Без них возможны дубли, пропуски или некорректные связи между сущностями. B: Доменные ограничения минимизируют количество аномальных, пустых или несогласованных значений (например, возраст = –5), что снижает трудозатраты

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			на очистку данных перед анализом. D: Согласованные справочники (например, «Москва», а не «Мск», «г. Москва» и т.п.) обеспечивают корректную группировку и кодирование категориальных переменных — необходимое условие для точного расчёта статистик и построения моделей. ***Задания комбинированного типа, предполагающие выбор нескольких ответов из предложенных*** При подготовке данных из реляционной базы для построения логистической регрессии в Python (sklearn.linear_model.LogisticRegression) какие из следующих практик проектирования БД являются целесообразными? A) Хранение бинарных исходов (например, «совершил покупку / не совершил») в виде BOOLEAN или TINYINT(1) B) Использование раздельных таблиц для событий и характеристик клиентов с чёткой связью через ID C) Кодирование категориальных признаков непосредственно в SQL-запросе с помощью CASE-выражений D) Отсутствие индексов по полям, используемым в WHERE и JOIN E) Обеспечение уникальности записей о клиентах в выборке через PRIMARY KEY Ответ: A, B, E Обоснование: A: Корректный тип данных для целевой переменной (0/1 или TRUE/FALSE) упрощает её использование в модели без дополнительного преобразования и снижает риск ошибок. B: Разделение сущностей (клиенты ↔ события) соответствует принципам нормализации и позволяет гибко агрегировать признаки (например, «число посещений», «средний чек»), сохраняя целостность данных. E: Уникальность строк в обучающей выборке (по ID клиента) предотвращает дублирование наблюдений, которое может исказить оценку коэффициентов модели и завысить её точность.

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			Задания открытого типа с развернутым ответом. Почему грамотное проектирование базы данных (включая нормализацию, использование внешних ключей и доменных ограничений) является необходимым условием для корректного решения задач статистического анализа в таких средах, как R или Python? Ответ: Грамотное проектирование базы данных обеспечивает целостность, согласованность и полноту исходных данных, что напрямую влияет на достоверность результатов статистического анализа. Нормализация устраняет избыточность и аномалии обновления, предотвращая дублирование информации и искажение частотных распределений. Внешние ключи гарантируют корректные связи между сущностями (например, «клиент — заказ»), что критично при объединении таблиц для формирования единой аналитической выборки. Доменные ограничения (NOT NULL, CHECK, ENUM) минимизируют количество пропущенных, ошибочных или несопоставимых значений, снижая трудозатраты на предварительную очистку данных в pandas или dplyr. Без этих мер даже самые продвинутые статистические пакеты будут работать с «грязными» или неполными данными, что приведёт к ошибочным выводам и невалидным моделям. ***Задания открытого типа с развернутым ответом.*** Как использование измерительных моделей (например, звёздной схемы) при проектировании базы данных упрощает и ускоряет решение задач описательной и инференциальной статистики с помощью программных пакетов, таких как statsmodels, scikit-learn или ggplot2? Ответ: Звёздная схема, состоящая из центральной таблицы фактов и окружающих её таблиц измерений (dimensions), специально оптимизирована для аналитических запросов. Она позволяет быстро агрегировать количественные показатели

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Тест по теме
			(например, выручка, количество транзакций) по нужным категориям (регион, время, продукт) без сложных многоуровневых JOIN-операций, характерных для полностью нормализованных OLTP-систем. Это упрощает экспорт «плоской» структурированной выборки, идеально подходящей для загрузки в DataFrame (Python) или data.frame (R). Благодаря чёткому разделению метрик и измерений, аналитик легко формирует признаки для описательной статистики (средние, дисперсии, распределения) и инференциальных моделей (регрессия, ANOVA). Кроме того, согласованные измерения (например, единый календарь) обеспечивают корректность временных и категориальных сравнений, что особенно важно при визуализации в ggplot2 или интерпретации коэффициентов в statsmodels.

Критерии оценивания тестового задания:

Оценка	Критерии оценивания
отлично	от 90 до 100 % правильно выполненных заданий
хорошо	от 70 до 89 % правильно выполненных заданий
удовлетворительно	от 50 до 69 % правильно выполненных заданий
неудовлетворительно	менее 50 % правильно выполненных заданий

4.2 ТИПОВЫЕ ПРАКТИЧЕСКИЕ ЗАДАНИЯ

Практические задания должны отражать умение обучающегося применять осваиваемую компетенцию в практических ситуациях и при решении производственных задач.

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Задания к практическому занятию
Тема 1. Визуализация данных с использованием пакетов статистического анализа.	**ПК-1**	**ИПК-1.2** **ИПК-1.3**	***Расчетная задача 1.*** Вы проектируете базу данных для хранения результатов опроса 1000 респондентов. Каждый респондент указывает: • возраст (целое число от 18 до 90), • пол («М» или «Ж»), • оценку удовлетворённости (от 1 до 5). Определите минимальный объём памяти (в байтах), необходимый для хранения одной записи, если использовать следующие типы данных: • TINYINT (1 байт) для возраста и оценки, • CHAR(1) (1 байт) для пола. Сколько всего памяти (в килобайтах) потребуется для хранения всех 1000 записей без учёта служебной информации СУБД? Решение: Одна запись: 1 (возраст) + 1 (пол) + 1 (оценка) = 3 байта 1000 записей: 1000 × 3 = 3000 байт ≈ 2.93 КБ ***Расчетная задача 2.*** При разработке API для передачи данных в статистическую систему вы решили использовать JSON-формат. Каждая запись содержит поля: "id" (целое число), "score" (вещественное число с точностью до двух знаков после запятой). Пример записи: {"id": 42, "score": 87.65} Оцените среднюю длину одной записи в символах (с пробелами и знаками препинания). Сколько таких записей можно передать за один вызов API, если ограничение на размер ответа — **1024 байта**, а кодировка — UTF-8 (1 символ = 1 байт)? **Решение:** Длина примерной записи: {"id": 42, "score": 87.65} → 28 символов = **28 байт** Максимум записей: ⌊1024 / 28⌋ = **36 записей**

<table>
<tr><th>Тема</th><th>Код и формулировка компетенции</th><th>Индикаторы достижения компетенции</th><th>Задания к практическому занятию</th></tr>
<tr><td></td><td></td><td></td><td>
Расчетная задача 3.
Для построения столбчатой диаграммы в ggplot2 (R) или seaborn (Python) требуется табличная структура данных, где каждая строка — отдельное наблюдение. У вас есть данные о продажах трёх товаров (A, B, C) в двух регионах (Север, Юг):
• Товар A: Север — 120, Юг — 80
• Товар B: Север — 90, Юг — 110
• Товар C: Север — 70, Юг — 130
Сколько строк должно быть в «длинном» (long) формате таблицы, пригодной для визуализации?
Ответ:
В длинном формате каждая комбинация «Товар–Регион» — отдельная строка. 3 товара × 2 региона = 6 наблюдений. Пример структуры:
<table>
<tr><th>товар</th><th>регион</th><th>продажи</th></tr>
<tr><td>A</td><td>Север</td><td>120</td></tr>
<tr><td>A</td><td>Юг</td><td>80</td></tr>
<tr><td>B</td><td>Север</td><td>90</td></tr>
<tr><td>B</td><td>Юг</td><td>110</td></tr>
<tr><td>C</td><td>Север</td><td>70</td></tr>
<tr><td>C</td><td>Юг</td><td>130</td></tr>
</table>
Эта структура корректно поддерживается функциями geom_bar() (ggplot2) и sns.barplot() (seaborn).

Расчетная задача 4.
Для построения тепловой карты корреляции в Python с помощью seaborn.heatmap() используется квадратная матрица корреляций размером n × n, где n — число числовых переменных.
Если у вас есть DataFrame с 5 числовыми столбцами (x1, x2, x3, x4, x5), сколько элементов будет содержать матрица корреляций, и сколько из них уникальных (без учёта диагонали и симметрии)?
</td></tr>
</table>

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Задания к практическому занятию
			Ответ: • Всего элементов: 25 • Уникальных недиагональных элементов: 10 Обоснование: Матрица 5×5 → $5^2 = 25$ элементов. Диагональ содержит 5 единиц (корреляция переменной с самой собой). Остальные 20 элементов симметричны относительно диагонали, поэтому уникальных пар: 20 / 2 = 10. Именно эти 10 значений определяют структуру тепловой карты, так как sns.heatmap() отображает всю матрицу, но содержательно важны только уникальные корреляции.
Тема 2. Решение задач статистического анализа с использованием программных пакетов.	**ПК-1**	ИПК-1.2 ИПК-1.3	***Расчетная задача 1*** Требуется разработать модуль для автоматического расчёта описательных статистик (среднее, медиана, стандартное отклонение) по числовым данным из CSV-файла. Какой подход к проектированию ПО следует применить, чтобы обеспечить повторное использование кода и лёгкую интеграцию с другими модулями? Ответ: Следует использовать объектно-ориентированный подход с выделением класса StatAnalyzer, инкапсулирующего методы загрузки данных и расчёта статистик. Для обеспечения повторного использования и расширяемости применяются принципы SOLID, а взаимодействие с файловой системой выносится в отдельный сервис. Реализация может быть выполнена на Python с использованием библиотек pandas и NumPy. ***Расчетная задача 2*** При анализе данных необходимо визуализировать распределение признака и проверить его нормальность с помощью критерия Шапиро–Уилка. Как организовать архитектуру программного модуля, чтобы отделить логику анализа от представления результатов? Ответ:

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Задания к практическому занятию
			Рекомендуется применить паттерн MVC (Model–View–Controller): • Model реализует статистические вычисления (например, через SciPy); • View отвечает за визуализацию (например, через Matplotlib или Seaborn); • Controller координирует взаимодействие между ними. • Такая архитектура упрощает тестирование, поддержку и замену компонентов без изменения остальной части системы. ***Расчетная задача 3.*** Из реляционной базы данных PostgreSQL таблицы sales (поля: id, product_id, amount, sale_date) требуется извлечь среднее значение продаж (amount) за последний квартал и рассчитать стандартное отклонение. Как это реализовать с использованием Python и библиотеки pandas? Ответ: python 1.import pandas as pd 2.import psycopg2 3.from sqlalchemy import create_engine 4. *5.# Подключение к БД* engine = create_engine('postgresql://user:password@localhost/dbname') 7. *8.# SQL-запрос за последний квартал (пример для Q4 2025)* 9.query = """ 10. SELECT amount 11. FROM sales 12. WHERE sale_date >= '2025-10-01' AND sale_date <= '2025-12-31'; 13.""" 14. 15.df = pd.read_sql(query, engine) 16. 17.mean_sales = df['amount'].mean()

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Задания к практическому занятию
			18.std_sales = df['amount'].std() 19. 20.print(f"Среднее: {mean_sales:.2f}, Стандартное отклонение: {std_sales:.2f}") Результат (пример): Среднее: 1245.60, Стандартное отклонение: 320.75 ***Расчетная задача 4*** Через REST API (например, https://api.example.com/data) доступны ежедневные данные о посещаемости сайта в формате JSON: Json [{"date": "2026-01-01", "visitors": 1200}, {"date": "2026-01-02", "visitors": 1350}, ...] Требуется загрузить данные за январь 2026 г., сохранить их в SQLite-таблицу traffic, а затем вычислить медиану и 95%-й перцентиль числа посетителей. Ответ: python 1.import requests 2.import sqlite3 3.import pandas as pd 4 *5.# Получение данных через API* 6.response = requests.get("https://api.example.com/data?month=2026-01") 7.data = response.json() 8. *9.# Сохранение в SQLite* 10conn = sqlite3.connect("analytics.db") 11.df = pd.DataFrame(data) 12.df.to_sql("traffic", conn, if_exists="replace", index=False) 13. *14.# Статистический анализ*

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Задания к практическому занятию
			15.median_vis = df['visitors'].median() 16.p95_vis = df['visitors'].quantile(0.95) 17. 18.print(f"Медиана: {median_vis}, 95%-й перцентиль: {p95_vis:.1f}") Результат (пример): Медиана: 1280, 95%-й перцентиль: 1840.0 ***Расчетная задача 5*** Необходимо разработать многократно используемый модуль на Python для расчёта корреляционной матрицы по числовым данным из CSV-файла. Как применить принципы проектирования ПО (например, инкапсуляцию и разделение ответственности), чтобы модуль легко тестировался и интегрировался в другие проекты? Решение: Применим объектно-ориентированное проектирование и выделим отдельный класс CorrelationAnalyzer, соблюдая принцип единственной ответственности: python 1.import pandas as pd 2. 3.class CorrelationAnalyzer: 4. def __init__(self, file_path: str): 5. self.data = pd.read_csv(file_path) 6. 7. def compute_correlation_matrix(self, method='pearson'): 8. numeric_data = self.data.select_dtypes(include='number') 9. return numeric_data.corr(method=method) 10. 11.# Пример использования

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Задания к практическому занятию
			12.analyzer = CorrelationAnalyzer('data.csv') 13.corr_matrix = analyzer.compute_correlation_matrix() 14. print(corr_matrix) Такой подход обеспечивает: • чёткое разделение логики загрузки данных и анализа, • возможность подмены источника данных (через внедрение зависимости), • простоту unit-тестирования метода compute_correlation_matrix. ***Расчетная задача 6*** Требуется создать консольное приложение, которое: • Загружает данные из файла, • Выполняет базовый статистический анализ (среднее, дисперсия, гистограмма), • Сохраняет результаты в отчёт. Как спроектировать архитектуру приложения, чтобы обеспечить масштабируемость и лёгкую замену компонентов (например, вывода в файл или на экран)? Решение: Используем паттерн «Стратегия» и модульную архитектуру: python 1.# analysis.py 2. import pandas as pd 3. import matplotlib.pyplot as plt 4. 5. class DataProcessor: 6. def __init__(self, loader, reporter): 7. self.loader = loader # стратегия загрузки 8. self.reporter = reporter # стратегия отчёта

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Задания к практическому занятию
			9. 10. def run(self, source): 11. df = self.loader.load(source) 12. stats = { 13. 'mean': df.mean(numeric_only=True).to_dict(), 14. 'var': df.var(numeric_only=True).to_dict() 15. } 16. self.reporter.generate(stats, df) 17 18.# loaders.py 19.class CSVLoader: 20. def load(self, path): 21. return pd.read_csv(path) 22. 23.# reporters.py 24.class FileReporter: 25. def generate(self, stats, df): 26. with open('report.txt', 'w') as f: Такая архитектура: • позволяет легко заменить формат входных данных (CSV → JSON → БД), • поддерживает разные типы отчётов (файл, email, веб-интерфейс), • соответствует принципам гибкого проектирования ПО (SOLID, loose coupling).

Критерии оценивания практических занятий:

Оценка	Критерии оценивания
отлично	Выставляется, если обучающийся умеет увязывать теорию с практикой (решает задачи, формулирует выводы, умеет пояснить полученные результаты), владеет понятийным аппаратом, полно и глубоко овладел материалом по заданной теме, обосновывает свои суждения и даёт правильные ответы на вопросы преподавателя

хорошо	Выставляется, если обучающийся умеет увязывать теорию с практикой (решает задачи и формулирует выводы, умеет пояснить полученные результаты), владеет понятийным аппаратом, полно и глубоко овладел материалом по заданной теме, но содержание ответов имеют некоторые неточности и требуют уточнения и комментария со стороны преподавателя
удовлетворительно	Выставляется, если обучающийся знает и понимает материал по заданной теме, но изложение неполное, непоследовательное, допускаются неточности в определении понятий, студент не может обосновать свои ответы на уточняющие вопросы преподавателя
неудовлетворительно	Выставляется, если обучающийся допускает ошибки в определении понятий, искажающие их смысл, беспорядочно и неуверенно излагает материал. Делает ошибки в ответах на уточняющие вопросы преподавателя

4.3. ТИПОВЫЕ ЗАДАНИЯ ДЛЯ КОНТРОЛЬНЫХ РАБОТ

Контрольные работы содержат несколько практических заданий по индивидуальным вариантам, в полном объеме охватывающих изученный материал по указанной теме. Выполнение контрольных работ позволяет определить результат освоения компетенций по дисциплине в рамках рассматриваемой темы, оцениваемый с помощью соответствующих индикаторов достижения компетенций.

Тема	**Код и формулировка компетенции**	**Индикаторы достижения компетенции**	**Типовой вариант контрольной работы**
Тема 1. Визуализация данных с использованием пакетов статистического анализа.	**ПК-1**	**ИПК-1.2**	***Контрольная работа:*** Компания ведёт учёт продаж в трёх таблицах реляционной базы данных: • products (product_id, product_name, category) • sales (sale_id, product_id, sale_date, quantity, revenue) • categories (category_id, category_name) • Спроектируйте ER-диаграмму (описательно, без графики): укажите сущности, атрибутes и связи между таблицами. • Напишите SQL-запрос, который получает суммарную выручку по категориям за последний месяц. • Опишите, как вы визуализируете эти данные в Python с помощью библиотеки matplotlib или seaborn, считывая результат запроса через pandas. Приведите фрагмент кода. Ответ

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Типовой вариант контрольной работы

			1. Описание ER-диаграммы:
			• Сущность products: содержит информацию о товарах; связан с categories через внешний ключ category → category_id (связь «многие к одному»).
			• Сущность sales: содержит записи о продажах; связан с products через product_id → product_id (связь «многие к одному»).
			• Сущность categories: справочник категорий; связана с products.
			
			2. SQL-запрос (предположим, текущая дата — январь 2026 г.):
			sql
			1.SELECT c.category_name, SUM(s.revenue) AS total_revenue
			2.FROM sales s
			3.JOIN products p ON s.product_id = p.product_id
			4.JOIN categories c ON p.category = c.category_id
			5.WHERE s.sale_date >= '2026-01-01' AND s.sale_date <= '2026-01-31'
			6.GROUP BY c.category_name;
			
			3. Визуализация в Python:
			
			python
			1.import pandas as pd
			2.import matplotlib.pyplot as plt
			3.from sqlalchemy import create_engine
			4.
			5.engine = create_engine('sqlite:///sales.db') # или другой URI
			6.query = "..." # запрос из п.2
			7.
			8.df = pd.read_sql(query, engine)
			9.plt.figure(figsize=(8, 5))
			10.plt.bar(df['category_name'], df['total_revenue'])
			11.plt.title('Выручка по категориям за январь 2026')
			12.plt.xlabel('Категория')

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Типовой вариант контрольной работы
			13.plt.ylabel('Выручка, руб.') 14.plt.xticks(rotation=45) 15.plt.tight_layout() 16.plt.show() Комментарий: Такой подход демонстрирует интеграцию грамотно спроектированной БД (нормализованной структуры), SQL-аналитики и визуализации — ключевые компетенции в анализе данных.
Тема 2. Решение задач статистического анализа с использованием программных пакетов.	**ПК-1**	**ИПК-1.2**	***Контрольная работа.*** Имеется файл exam_scores.csv со столбцами: • student_id — идентификатор студента, • group — учебная группа (A, B или C), • score — балл за экзамен (от 0 до 100). Выполните следующие действия с использованием Python и библиотек pandas, numpy, scipy и/или matplotlib: 1. Загрузите данные и проверьте наличие пропущенных значений. 2. Рассчитайте средний балл и стандартное отклонение по каждой группе. 3. Проверьте гипотезу о том, что средние баллы в группах A и B не различаются статистически значимо (используйте t-критерий Стьюдента при уровне значимости $\alpha = 0.05$). 4. Постройте boxplot для сравнения распределений баллов по трём группам. Ответ: python 1.import pandas as pd 2.import numpy as np 3.from scipy import stats 4.import matplotlib.pyplot as plt 5. 6. # 1. Загрузка и проверка пропусков 7. df = pd.read_csv('exam_scores.csv')

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Типовой вариант контрольной работы
			8. print("Пропущенные значения:", df.isnull().sum().sum()) 9. 10. # 2. Описательная статистика по группам 11.group_stats = df.groupby('group')['score'].agg(['mean', 'std']) 12. print("\nСреднее и стандартное отклонение по группам:") 13. print(group_stats) 14. 15. # 3. t-тест для групп A и B 16. group_A = df[df['group'] == 'A']['score'] 17. group_B = df[df['group'] == 'B']['score'] 18. t_stat, p_value = stats.ttest_ind(group_A, group_B, equal_var=False) 19. 20. print(f"\nt-статистика: {t_stat:.3f}, p-значение: {p_value:.3f}") 21. if p_value < 0.05: 22. print("Различия статистически значимы (отвергаем H0).") 23.else: 24. print("Нет оснований отвергать H0 (различия не значимы).") 25 26 # 4. Визуализация 27. plt.figure(figsize=(6, 4)) 28. df.boxplot(column='score', by='group', ax=plt.gca()) 29. plt.title('Распределение баллов по группам') 30. plt.suptitle('') 31. plt.ylabel('Балл') 32. plt.show() Пример вывода 1.Пропущенные значения: 0 2. 3.Среднее и стандартное отклонение по группам: 4. mean std

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Типовой вариант контрольной работы
			5. group 6. A 78.4 9.2 7. B 75.1 10.5 8. C 82.0 8.7 9. 10. t-статистика: 1.420, p-значение: 0.162 11. Нет оснований отвергать H0 (различия не значимы). Комментарий: Решение демонстрирует комплексное применение программных пакетов для загрузки данных, описательного анализа, проверки статистических гипотез и визуализации — ключевые навыки при решении задач статистического анализа в прикладной среде.

Критерии оценивания контрольной работы:

Оценка	Критерии оценивания
отлично	Выставляется, если обучающийся умеет увязывать теорию с практикой (решает задачи, формулирует выводы, умеет пояснить полученные результаты), владеет понятийным аппаратом, полно и глубоко овладел материалом по заданной теме, обосновывает свои суждения и даёт правильные ответы на вопросы преподавателя
хорошо	Выставляется, если обучающийся умеет увязывать теорию с практикой (решает задачи и формулирует выводы, умеет пояснить полученные результаты), владеет понятийным аппаратом, полно и глубоко овладел материалом по заданной теме, но содержание ответов имеют некоторые неточности и требуют уточнения и комментария со стороны преподавателя.
удовлетворительно	Выставляется, если обучающийся знает и понимает материал по заданной теме, но изложение неполное, непоследовательное, допускаются неточности в определении понятий, студент не может обосновать свои ответы на уточняющие вопросы преподавателя
неудовлетворительно	Выставляется, если обучающийся допускает ошибки в определении понятий, искажающие их смысл, беспорядочно и неуверенно излагает материал. Делает ошибки в ответах на уточняющие вопросы преподавателя

4.4. ТИПОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Самостоятельная работа включает в себя проработку теоретического материала, изучение рекомендуемой литературы, выполнение практико-ориентированных заданий (заполнение таблиц, проведение сравнительного анализа, составление схем и др.), решение практических задач, создание презентаций, написание рефератов, подборка нормативного и иного материала и выполнение других заданий.

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Типовое задание для самостоятельной работы
Тема 1. Визуализация данных с использованием пакетов статистического анализа.	ПК-1. Способен применять методы и средства проектирования компьютерного программного обеспечения	**ИПК-1.1** **ИПК-1.2**	***Решение практических задач*** Даны данные о продажах за 6 месяцев: Python 1.months = ['Янв', 'Фев', 'Мар', 'Апр', 'Май', 'Июн'] 2.sales = [120, 135, 150, 140, 160, 170] Постройте линейный график зависимости продаж от месяца с помощью библиотеки matplotlib. Добавьте заголовок «Продажи по месяцам» и подпишите оси. Ответ: Python 1.import matplotlib.pyplot as plt 2. 3.months = ['Янв', 'Фев', 'Мар', 'Апр', 'Май', 'Июн'] 4.sales = [120, 135, 150, 140, 160, 170] 5. 6.plt.plot(months, sales, marker='o') 7.plt.title('Продажи по месяцам') 8.plt.xlabel('Месяц') 9. plt.ylabel('Продажи') 10.plt.grid(True) 11. plt.show() График покажет динамику роста продаж за полгода.
Тема 2. Решение задач статистическо	ПК-1. Способен применять методы и средства	**ИПК-1.1** **ИПК-1.2**	***Задание на проведение сравнительного анализа*** Даны два небольших набора данных — время (в секундах), за которое две группы пользователей выполнили одну и ту же задачу:

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Типовое задание для самостоятельной работы
го анализа с использовани ем программных пакетов.	проектирования компьютерного программного обеспечения		• Группа A (опытные): [45, 48, 50, 42, 47] • Группа B (новички): [60, 65, 58, 70, 62] С помощью Python (scipy и matplotlib) выполните следующее: 1. Рассчитайте среднее время в каждой группе. 2. Постройте столбчатую диаграмму для сравнения средних. ***3.*** Проведите t-тест и определите, отличаются ли группы статистически значимо (α = 0.05). Ответ: Python 1. import matplotlib.pyplot as plt 2.from scipy import stats 3.import numpy as np 4 5 # Данные 6 A = [45, 48, 50, 42, 47] 7 B = [60, 65, 58, 70, 62] 8 9 # 1. Средние значения 10 mean_A = np.mean(A) 11 mean_B = np.mean(B) 12 print(f"Среднее A: {mean_A:.1f}, Среднее B: {mean_B:.1f}") 13 14 # 2. Столбчатая диаграмма 15 plt.bar(['Группа A', 'Группа B'], [mean_A, mean_B]) 16 plt.ylabel('Среднее время (сек)') 17 plt.title('Сравнение времени выполнения задачи') 18 plt.show() 19 20 # 3. t-тест

Тема	Код и формулировка компетенции	Индикаторы достижения компетенции	Типовое задание для самостоятельной работы
			21 t_stat, p_val = stats.ttest_ind(A, B) 22 print(f"p-значение: {p_val:.4f}") 23 if p_val < 0.05: 24 print("Различия статистически значимы.") 25 else: 26 print("Различия не значимы.") Выводы: 1. Среднее A: 46.4, Среднее B: 63.0 2. p-значение: 0.0021 3. Различия статистически значимы.

Критерии оценивания самостоятельной работы:

Оценка	Критерии оценивания
отлично	выставляется если работа носит научно-исследовательский характер, проанализирован и сделан сравнительный анализ нескольких литературных источников, приведены примеры
хорошо	выставляется если проанализирован и сделан сравнительный анализ нескольких литературных источников, приведены примеры
удовлетворительно	выставляется если проведен сравнительный анализ научно-методической литературы, приведены примеры
неудовлетворительно	выставляется если работа прошла проверку на антиплагиат и соответствует требованиям оформления

4.5. ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ ДЛЯ КУРСОВЫХ РАБОТ

Курсовая работа позволяет выявить степень владения базовыми знаниями, умениями и навыками, необходимыми для обучения, и определить уровень владения новым материалом.

Примерные индивидуальные задания (темы) для курсовых работ:

Курсовая работа не предусмотрена учебным планом

Критерии оценивания курсовой работы:

Оценка	Критерии оценивания
отлично	Содержание курсовой работы полностью соответствует заданию, содержащемуся в методических указаниях, и плану. Представлены результаты структурированного и логически последовательного обзора литературных и иных

	источников по теме исследования. Структура курсовой работы логически и методически выдержана. Верно определены исходные данные для расчетов. Все аналитические расчеты выполнены верно, корректно применены методы экономического анализа, не нарушена методика анализа предмета исследования. Все выводы и предложения убедительно аргументированы. При защите курсовой работы обучающийся правильно и уверенно отвечает на вопросы преподавателя, демонстрирует глубокое знание теоретического материала, способен аргументировать собственные утверждения и выводы
хорошо	Содержание курсовой работы полностью соответствует заданию, содержащемуся в методических указаниях, и плану. Представлены результаты структурированного и логически последовательного обзора литературных и иных источников по теме исследования. Структура курсовой работы логически и методически выдержана. Верно определены исходные данные для расчетов. В расчетах допускаются незначительные (не искажающие общего итога оценки) погрешности/ошибки. Большинство выводов и предложений аргументировано, корректно применены методы экономического анализа, не нарушена методика анализа предмета исследования. Оформление курсовой работы и полученные результаты в целом отвечают требованиям, изложенным в методических указаниях. Имеются одна-две несущественные ошибки в использовании терминов, в построенных диаграммах и схемах, в оформлении таблиц. Наличествует незначительное количество грамматических и/или стилистических ошибок. При защите курсовой работы обучающийся правильно и уверенно отвечает на большинство вопросов преподавателя, демонстрирует хорошее знание теоретического материала, но не всегда способен аргументировать собственные утверждения и выводы. При наводящих вопросах преподавателя исправляет ошибки в ответе
удовлетворительно	Содержание курсовой работы полностью соответствует заданию, содержащемуся в методических указаниях, и плану. Результаты обзора литературных и иных источников представлены недостаточно полно, недостаточно логично и последовательно. Верно определены исходные данные для расчетов, но имеются грубые ошибки в расчетах. Аргументация выводов и предложений слабая или отсутствует. Экономические выводы носят констатирующий (описательный) характер. Имеются одно-два существенных отклонений от требований в оформлении курсовой работы. Полученные результаты в целом отвечают требованиям, изложенным в методических указаниях. Имеются одна-две существенных ошибки в использовании терминов, в построенных диаграммах и схемах. Много грамматических и/или стилистических ошибок. При защите курсовой работы обучающийся допускает грубые ошибки при ответах на вопросы преподавателя, демонстрирует слабое знание теоретического материала, в большинстве случаев не способен уверенно аргументировать собственные утверждения и выводы
неудовлетворительно	Содержание курсовой работы не соответствует заданию, содержащемуся в методических указаниях, и плану. Неверно определены исходные данные для расчетов, неверно и не корректно применены методы экономического анализа. Экономические выводы содержат неверную экономическую оценку. Имеются более двух существенных отклонений от требований в оформлении курсовой работы. Большое количество существенных ошибок по сути работы, много грамматических и стилистических ошибок и др. Полученные результаты не

	отвечают требованиям, изложенным в методических указаниях. При защите курсовой работы обучающийся демонстрирует слабое понимание программного материала, студент не может защитить свои решения, допускает грубые фактические ошибки при ответах на поставленные вопросы или вовсе не отвечает на них. Курсовая работа не представлена преподавателю. Обучающийся не явился на защиту курсовой работы

5. МАТЕРИАЛЫ ДЛЯ ОЦЕНКИ РЕЗУЛЬТАТОВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Промежуточная аттестация проводится в форме зачет.

1.1. Вопросы к зачету в семестре 7

Недостаточная точность результатов статистического анализа из-за ограничений моделей или алгоритмов
2. Сложность интерпретации статистических результатов и их применения на практике
3. Ошибки при обработке и анализе больших объемов данных в статистическом программном обеспечении
4. Проблемы с выбором подходящих статистических методов и их правильной настройкой
5. Недостаток гибкости статистического программного обеспечения для адаптации к разнообразным исследовательским задачам
6. Проблемы с обучением и переобучением моделей в статистическом программном обеспечении
7. Ограничения в возможностях визуализации статистических данных и результатов анализа
8. Проблемы с качеством и достоверностью статистических данных, вводимых в программное обеспечение
9. Недостаточные средства для работы с пропущенными данными и анализом выбросов в статистическом программном обеспечении
10. Ограничения в скорости выполнения статистических вычислений в программном обеспечении
11. Проблемы с масштабированием статистического анализа на большие объемы данных
12. Недостаток поддержки высокопроизводительных вычислений и параллельных вычислений в статистическом программном обеспечении
13. Проблемы с согласованием результатов статистического анализа, полученных из различных источников
14. Ограничения в возможностях интеграции статистического программного обеспечения с другими информационными системами
15. Недостаточное управление версиями и контроль изменений в статистических моделях и анализе
16. Проблемы с защитой и сохранностью данных при работе со статистическим программным обеспечением
17. Ограниченная поддержка различных языков программирования и форматов данных в статистическом программном обеспечении
18. Проблемы с разработкой и поддержкой пользовательских расширений и плагинов для статистического программного обеспечения
19. Недостаток доступности и качества обучающих материалов и документации по использованию статистического программного обеспечения
20. Проблемы с интерфейсом и удобством использования статистического программного обеспечения для различных категорий пользователей.

Критерии оценивания экзамена:

Оценка	Критерии оценивания
отлично	Выставляется, если обучающимся правильно и полностью раскрыто содержание материала в пределах программы,

	чётко и правильно даны определения и раскрыто содержание понятий, точно использованы научные и технические термины, в ответе использованы ранее приобретённые теоретические знания, сделаны необходимые выводы и обобщения
хорошо	Выставляется, если обучающимся раскрыто основное содержание материала в пределах программы, даны определения и раскрыто содержание понятий, в ответе использованы ранее приобретённые теоретические знания, сделаны необходимые выводы и обобщения, но присутствуют незначительные нарушения в последовательности изложения, имеются одна-две неточности в содержании ответа
удовлетворительно	Выставляется, если обучающимся содержание учебного материала изложено фрагментарно, не всегда последовательно, не даны определения, не раскрыто содержание понятий, или они изложены с ошибками, допускаются ошибки и неточности в использовании научной терминологии, отсутствуют выводы и обобщения из предыдущего материала, или возможны ошибки в их изложении
неудовлетворительно	Выставляется, если обучающимся основное содержание учебного материала не раскрыто, не даются ответы на основные вопросы, допускаются грубые ошибки в определении понятий, в использовании терминологии, отсутствуют выводы и обобщения

6. МАТЕРИАЛЫ ДЛЯ ДИАГНОСТИЧЕСКОЙ РАБОТЫ

Задания для диагностической работы должны обеспечивать оценку полностью или частично сформированных компетенций. Каждое задание должно быть привязано к тому или иному индикатору сформированности компетенций.

При формировании заданий для диагностической работы необходимо использовать тестовые задания следующих типов:

Тип задания 1. Задания закрытого типа на установление соответствия.

Тип задания 2. Задания закрытого типа на установление последовательности.

Тип задания 3. Задания комбинированного типа, предполагающие выбор одного правильного ответа из предложенных

Тип задания 4. Задания комбинированного типа, предполагающие выбор нескольких ответов из предложенных

Тип задания 5. Задания открытого типа с развернутым ответом.

Все типы заданий должны быть представлены не менее одного раза.

№ п/п	**Тема занятия**	**Код компетенции**	**Индикатор**	**Тип задания**	**Задание**		Уровень освоения компетенции
					Вариант 1	**Вариант 2**	

<table>
<tr>
<td>1.</td>
<td>Тема 1. Визуализация данных с использованием пакетов статистического анализа.</td>
<td>ПК-1</td>
<td>ИПК-1.1</td>
<td>Задания закрытого типа на установление соответствия.</td>
<td>1. Знание темы визуализация данных с использованием пакетов статистического анализа необходимо для эффективного представления сложных количественных результатов в наглядной и интерпретируемой форме, что особенно важно при принятии обоснованных управленческих и аналитических решений. Установите соответствие между типом визуализации и её наиболее подходящим применением в рамках методологии проектирования аналитических дашбордов.
<table>
<tr><td></td><td>Тип визуализации</td><td></td><td>Применение</td></tr>
<tr><td>А</td><td>Гистограмма</td><td>1</td><td>Отображение распределения категориальных данных</td></tr>
<tr><td>Б</td><td>Круговая диаграмма</td><td>2</td><td>Показ соотношения частей к целому при небольшом числе категорий</td></tr>
<tr><td>В</td><td>Диаграмма рассеяни</td><td>3</td><td>Анализ формы распределени</td></tr>
</table>
</td>
<td>1. Знание методов и средств проектирования баз данных в сочетании с навыками визуализации с использованием пакетов статистического анализа необходимо для обеспечения целостности, эффективного извлечения и наглядного представления данных на всех этапах аналитического цикла — от хранения до принятия решений. Установите соответствие между пакетом статистического анализа и функцией, используемой для создания графиков в соответствии с принципами визуального проектирования.
<table>
<tr><td></td><td>Пакет / библиотека</td><td></td><td>Функция / инструмент визуализации</td></tr>
<tr><td>А</td><td>R (ggplot2)</td><td>1</td><td>sns.heatmap()</td></tr>
<tr><td>Б</td><td>Python (Matplotlib)</td><td>2</td><td>geom_bar()</td></tr>
<tr><td>В</td><td>Python (Seaborn)</td><td>3</td><td>plt.scatter()</td></tr>
<tr><td>Г</td><td>R (base graphics)</td><td>4</td><td>barplot()</td></tr>
</table>
Ответ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td>2</td><td>3</td><td>1</td><td>4</td></tr>
</table>
</td>
<td>Базовый (1-3 минуты)</td>
</tr>
</table>

<table>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>
<table>
<tr><td></td><td>я (scatter plot)</td><td></td><td>я количественной переменной</td></tr>
<tr><td>Г</td><td>Столбчатая диаграмма</td><td>4</td><td>Выявление корреляции между двумя числовыми переменными</td></tr>
</table>
Ответ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td><td>А</td></tr>
<tr><td>3</td><td>2</td><td>4</td><td>1</td><td>3</td></tr>
</table>
</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Тема 1. Визуализация данных с использованием пакетов статистического анализа.</td>
<td>ПК-1</td>
<td>ИПК-1.1</td>
<td>Задания закрытого типа на установление последовательности.</td>
<td>2. Установите правильную последовательность этапов работы с данными — от проектирования базы данных до визуализации результатов анализа с использованием статистических пакетов
1.Определение сущностей и связей, создание ER-диаграммы
2.Нормализация таблиц и реализация реляционной структуры в СУБД
3.Извлечение и агрегация данных с помощью SQL-запросов
4.Загрузка данных в среду анализа (например, R или Python)
5.Построение графиков (например, с помощью ggplot2 или seaborn) для</td>
<td>2. Расположите в логической последовательности действия, которые выполняет аналитик при подготовке отчёта на основе данных из спроектированной базы, используя пакеты статистического анализа.
1.Проектирование логической модели базы данных с учётом будущих аналитических запросов
2.Реализация физической модели базы данных и наполнение её данными
3.Экспорт выборки данных в формат, совместимый со статистическим пакетом (CSV, Parquet и др.)
4.Предварительная обработка и трансформация данных в R/Python
5.Визуализация ключевых метрик с</td>
<td>Базовый (1-3 минуты)</td>
</tr>
</table>

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>интерпретации результатов
Ответ:
<table><tr><td></td><td></td><td></td><td></td><td></td></tr></table>Ключ:
<table><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table></td><td>применением принципов эффективного графического дизайна
Ответ:
<table><tr><td></td><td></td><td></td><td></td><td></td></tr></table>Ключ:
<table><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table></td><td></td></tr>
<tr><td>3</td><td>Тема 1. Визуализация данных с использованием пакетов статистического анализа.</td><td>ПК-1</td><td>ИПК-1.1</td><td>Задания комбинированного типа, предполагающие выбор одного правильного ответа из предложенных</td><td>3. При разработке системы аналитики продаж компания планирует визуализировать динамику выручки по регионам и категориям товаров за последние три года. Какой подход к проектированию базы данных наиболее рационален для обеспечения эффективной последующей визуализации в R или Python?
A) Хранение всех данных в одной неструктурированной таблице без нормализации
B) Использование денормализованной «плоской» таблицы, специально подготовленной для аналитических запросов (например, в формате звездной схемы)
C) Полная нормализация до пятой нормальной формы с множеством связанных таблиц
D) Отказ от реляционной модели в пользу хранения данных исключительно в JSON-файлах
Ответ: B</td><td>3. Аналитик получил задачу визуализировать частоту обращений клиентов по типам проблем, с разбивкой по месяцам и каналам поддержки. На каком этапе проектирования базы данных закладываются предпосылки для корректной и гибкой визуализации таких данных?
A) На этапе выбора СУБД (например, MySQL vs PostgreSQL)
B) На этапе логического проектирования — при определении сущностей, атрибутов и связей
C) На этапе физического размещения базы данных на сервере
D) На этапе написания скриптов визуализации в R или Python
Ответ: B
Обоснование: Именно на этапе логического проектирования определяется структура данных: выделяются сущности (например, «Обращение», «Клиент», «Канал поддержки»), их атрибуты (включая дату, тип проблемы) и связи. Грамотно спроектированная модель</td><td>Повышенный (3-5 минут)</td></tr>
</table>

					Обоснование: Звездная схема (денормализованная структура с фактами и измерениями) оптимизирует выполнение агрегирующих запросов, упрощает извлечение данных и минимизирует сложность JOIN-операций, что позволяет быстрее загружать согласованные наборы данных в среды визуализации (например, pandas + seaborn или ggplot2), обеспечивая корректность и производительность аналитических графиков.	обеспечивает полноту, непротиворечивость и удобство извлечения данных, что напрямую влияет на возможность построения точных, детализированных и интерпретируемых визуализаций с помощью статистических пакетов.	
4	Тема 1. Визуализация данных с использованием пакетов статистического анализа.	**ПК-1**	**ИПК-1.1**	*Задания комбинированного типа, предполагающие выбор нескольких ответов из предложенных*	4. Какие из перечисленных характеристик проектирования базы данных напрямую влияют на качество и эффективность последующей визуализации данных с использованием пакетов статистического анализа (например, ggplot2, seaborn, matplotlib)? A) Наличие четко определённых внешних ключей между таблицами B) Использование осмысленных и согласованных названий столбцов C) Применение сжатия данных на уровне СУБД D) Нормализация или денормализация структуры в зависимости от целей анализа E) Хранение всех данных в формате BLOB	4. При подготовке данных из реляционной базы для визуализации временных трендов (например, ежедневных продаж) какие из следующих действий на этапе проектирования или подготовки данных являются целесообразными? A) Создание отдельной таблицы-измерения «Дата» с атрибутами (день недели, месяц, квартал, праздник/нет) B) Хранение даты как текстовой строки в произвольном формате (например, «22 января 2025») C) Обеспечение полноты временного ряда (включая дни без продаж) через LEFT JOIN с календарной таблицей D) Использование триггеров для автоматической визуализации новых записей	**Повышенный (3-5 минут)**

					Правильные ответы: A, B, D Обоснование: A: Четкие внешние ключи обеспечивают корректные связи между сущностями (например, «Клиент» — «Заказ»), что позволяет точно агрегировать и объединять данные перед визуализацией. B: Согласованные и понятные названия столбцов упрощают загрузку данных в аналитические среды (R/Python) и снижают риск ошибок при построении графиков. D: Выбор между нормализацией (для целостности) и денормализацией (для скорости аналитических запросов, например, в звёздной схеме) напрямую влияет на производительность и удобство извлечения данных, используемых в визуализации.	E) Применение типов данных DATE/DATETIME для временных меток Ответ: A, C, E Обоснование: A: Таблица-измерение «Дата» (как часть измерительной модели) упрощает группировку и фильтрацию по временным признакам, что критично для построения осмысленных временных графиков. C: Заполнение пропусков в временных рядах (например, дней без продаж) предотвращает искажение визуализации — линейные графики будут отображать реальную динамику, а не «скачки» из-за отсутствующих точек. E: Использование стандартных временных типов (DATE/DATETIME) обеспечивает корректную сортировку, агрегацию и совместимость с функциями визуализации в R/Python (например, pd.to_datetime()).	
5	Тема 1 Визуализация данных с использованием пакетов статисти	**ПК-1**	**ИПК-1.1**	*Задания открытого типа с развернутым ответом.*	5. Почему при проектировании базы данных важно учитывать будущие задачи визуализации, и как ошибки на этапе проектирования могут повлиять на качество графиков, создаваемых в R или Python? Ответ: При проектировании базы данных важно заранее учитывать	5. Как использование измерительных моделей (например, звёздной схемы) при проектировании базы данных упрощает процесс визуализации данных с помощью пакетов статистического анализа, таких как ggplot2 или seaborn? Ответ: Звёздная схема, состоящая из	**Высокий (5-10 минут)**

	ческого анализа.				аналитические и визуализационные цели, поскольку от структуры хранения зависит полнота, согласованность и доступность данных. Например, отсутствие нормализованной таблицы «Время» или некорректное хранение даты как текста затрудняет построение временных рядов. Аналогично, дублирование значений, отсутствие внешних ключей или несогласованные категории (например, «Москва» и «г. Москва») приводят к искажению агрегированных показателей и, как следствие, к ошибочным выводам на графиках. Таким образом, грамотное проектирование — основа достоверной и эффективной визуализации.	центральной таблицы фактов и окружающих её таблиц измерений (dimensions), оптимизирует выполнение аналитических запросов: она минимизирует сложность JOIN-операций и обеспечивает быстрое извлечение агрегированных данных с четкой семантикой (например, продажи по регионам, категориям, периодам). Это позволяет легко загружать структурированные, готовые к анализу датафреймы в R или Python, где пакеты вроде ggplot2 или seaborn требуют именно «плоских» табличных данных с ясными переменными. Благодаря этому аналитик тратит меньше времени на очистку и преобразование, а визуализации получаются точными, гибкими и интерпретируемыми.	
6	Тема 2 Решение задач статистического анализа с использованием программных пакетов.	**ПК-1**	**ИПК-1.1**	*Задания закрытого типа на установление соответствия*	6. Знание этапов проектирования базы данных необходимо, поскольку оно обеспечивает структурную целостность, согласованность и доступность данных, что напрямую определяет качество, точность и эффективность решения задач статистического анализа с использованием программных пакетов, таких как R и Python. Установите соответствие между этапом проектирования базы данных и его вкладом в успешное	6. Знание особенностей проектирования базы данных необходимо, поскольку оно позволяет формировать структурированные, непротиворечивые и эффективно извлекаемые наборы данных, что напрямую повышает точность, производительность и интерпретируемость результатов статистического анализа, выполняемого с помощью программных пакетов, таких как R и	**Базовый (1-3 минуты)**

					решение задач статистического анализа с использованием программных пакетов (R, Python и др.). |	Этап проектирования БД		Вклад в статистический анализ
---	---	---	---
А	Концептуальное проектирование (ER-модель)	1	Обеспечивает целостность и однозначную интерпретацию переменных, используемых в моделях
Б	Логическое проектирование (реляционная схема)	2	Позволяет точно определить сущности и связи, необходимые для формирования аналитических выборок
В	Физическое проектирование (индексы, типы данных)	3	Ускоряет извлечение больших объёмов данных, что критично при загрузке в pandas или R
Г	Нормал	4	Устраняет		Python. Установите соответствие между особенностью проектирования базы данных и преимуществом, которое она даёт при решении задач статистического анализа с помощью программных пакетов. |	Особенность проектирования БД		Преимущество для статистического анализа
---	---	---	---
А	Использование типов данных DATE/DATETIME	1	Позволяет корректно строить временные ряды и применять функции временной агрегации
Б	Создание внешних ключей между таблицами	2	Обеспечивает возможность безопасного объединения (JOIN) данных из разных сущностей
В	Применение доменных ограничений (CHECK, ENUM)	3	Снижает количество пропущенных или аномальных значений, упрощая предобработк	

<table>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>
<table>
<tr><td></td><td>изация таблиц</td><td></td><td>избыточность и аномалии, повышая качество данных для последующего анализа</td></tr>
</table>
Ответ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td>2</td><td>1</td><td>3</td><td>4</td></tr>
</table>
</td>
<td>
<table>
<tr><td></td><td></td><td></td><td>у</td></tr>
<tr><td>Г</td><td>Денормализация для аналитических нужд</td><td>4</td><td>Упрощает экспорт «плоской» таблицы, готовой к загрузке в pandas или data.frame</td></tr>
</table>
Ответ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>
Ключ:
<table>
<tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr>
<tr><td>1</td><td>2</td><td>в</td><td>4</td></tr>
</table>
</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Тема 2 Решение задач статистического анализа с использованием программных пакетов.</td>
<td>ПК-1</td>
<td>ИПК-1.1</td>
<td>Задания закрытого типа на установление последовательности</td>
<td>7. Установите правильную последовательность этапов работы с данными — от проектирования базы данных до выполнения статистического анализа с использованием программных пакетов (например, R или Python).
1.Определение предметной области и выявление ключевых сущностей и связей (построение ER-модели)
2.Разработка реляционной схемы: нормализация, определение первичных и внешних ключей
3.Реализация физической модели БД в СУБД и наполнение данными
4.Извлечение релевантной выборки с помощью SQL-запросов
5.Загрузка данных в среду анализа (например, в pandas DataFrame или R data.frame) и выполнение</td>
<td>7. Расположите в логическом порядке действия, обеспечивающие корректное решение задачи регрессионного анализа на основе данных из корпоративной базы.
1.Проектирование таблицы фактов и измерений с учётом будущих аналитических признаков (например, «доход», «регион», «дата»)
2.Обеспечение целостности данных через ограничения (NOT NULL, FOREIGN KEY, CHECK)
3.Экспорт подготовленного набора данных в формат CSV или прямое подключение к БД из Python/R
4.Предварительная обработка данных (обработка пропусков, кодирование категорий) в среде статистического анализа
5.Построение и интерпретация</td>
<td>Базовый (1-3 минуты)</td>
</tr>
</table>

					статистических расчётов Ответ: 1 | 2 | 3 | 4 | 5	регрессионной модели с использованием пакетов (statsmodels, scikit-learn, lm() в R) Ответ: 1 | 2 | 3 | 4 | 5	
8	Тема 2 Решение задач статистического анализа с использованием программных пакетов.	**ПК-1**	**ИПК-1.1**	*Задания комбинированного типа, предполагающие выбор одного правильного ответа из предложенных*	8. Аналитик планирует провести кластерный анализ клиентов на основе их покупательского поведения. Какой подход к проектированию базы данных наиболее рационален для обеспечения корректности и эффективности этого анализа при последующей работе в Python (с использованием pandas и scikit-learn)? A) Хранение всех событий в одной неструктурированной таблице без внешних ключей B) Использование полностью нормализованной модели до пятой нормальной формы C) Применение денормализованной структуры (например, звёздной схемы) с центральной таблицей фактов и измерениями «Клиент», «Товар», «Время» D) Отказ от реляционной модели в пользу хранения данных только в JSON-файлах Выберите один правильный. Ответ: C Обоснование: Звёздная схема	8. При решении задачи линейной регрессии в R (lm()) аналитику необходимо использовать данные о продажах, клиентах и рекламных кампаниях. На каком этапе проектирования базы данных закладываются основы для корректного формирования регрессионной выборки? A) На этапе выбора аппаратного обеспечения сервера СУБД B) На этапе логического проектирования — при определении сущностей, атрибутов и связей C) На этапе написания скрипта визуализации в ggplot2 D) На этапе настройки резервного копирования базы данных Выберите один правильный ответ. Ответ: B Обоснование: Именно на этапе логического проектирования определяется, какие переменные (атрибуты) будут доступны для анализа, как они связаны между собой и как будут объединяться (через внешние ключи). Например, если связь между «продажей» и	**Повышенный (3-5 минут)**

					обеспечивает баланс между целостностью данных и производительностью аналитических запросов. Таблица фактов содержит количественные метрики (например, сумма покупки, частота), а измерения — качественные признаки (регион, возраст, категория товара). Это позволяет легко агрегировать данные по клиентам и формировать готовый набор признаков для кластеризации в pandas. В отличие от полной нормализации (B), которая требует множества JOIN-запросов, или неструктурированного хранения (A, D), звёздная схема упрощает экспорт «плоской» таблицы, необходимой для алгоритмов машинного обучения.	«рекламной кампанией» не предусмотрена, соответствующий предиктор окажется недоступен, что исказит модель. Грамотная логическая модель гарантирует полноту, непротиворечивость и однозначную интерпретацию признаков — ключевых условий построения валидной регрессионной модели.	
9	Тема 2 Решение задач статистического анализа с использованием программных пакетов.	**ПК-1**	**ИПК-1.1**	*Задания комбинированного типа, предполагающие выбор нескольких ответ*	9. Какие из перечисленных элементов проектирования базы данных напрямую способствуют повышению качества и эффективности решения задач статистического анализа (например, в R с помощью dplyr/lm() или в Python с помощью pandas/statsmodels)? A) Использование первичных и внешних ключей B) Применение доменных ограничений (CHECK, ENUM, NOT	9. При подготовке данных из реляционной базы для построения логистической регрессии в Python (sklearn.linear_model.LogisticRegression) какие из следующих практик проектирования БД являются целесообразными? A) Хранение бинарных исходов (например, «совершил покупку / не совершил») в виде BOOLEAN или TINYINT(1) B) Использование раздельных таблиц для событий и характеристик	**Повышенный (3-5 минут)**

				ов из предложенных	NULL) C) Хранение всех данных в одном CSV-файле без использования СУБД D) Наличие согласованных справочников (например, единый список регионов) E) Отказ от нормализации в пользу полной денормализации Ответ: A, B, D Обоснование: A: Первичные и внешние ключи обеспечивают ссылочную целостность, что критично при объединении таблиц (JOIN) для формирования полной аналитической выборки. Без них возможны дубли, пропуски или некорректные связи между сущностями. B: Доменные ограничения минимизируют количество аномальных, пустых или несогласованных значений (например, возраст = –5), что снижает трудозатраты на очистку данных перед анализом. D: Согласованные справочники (например, «Москва», а не «Мск», «г. Москва» и т.п.) обеспечивают корректную группировку и кодирование категориальных переменных — необходимое условие для точного расчёта статистик и построения моделей.	клиентов с чёткой связью через ID C) Кодирование категориальных признаков непосредственно в SQL-запросе с помощью CASE-выражений D) Отсутствие индексов по полям, используемым в WHERE и JOIN E) Обеспечение уникальности записей о клиентах в выборке через PRIMARY KEY Ответ: A, B, E Обоснование: A: Корректный тип данных для целевой переменной (0/1 или TRUE/FALSE) упрощает её использование в модели без дополнительного преобразования и снижает риск ошибок. B: Разделение сущностей (клиенты ↔ события) соответствует принципам нормализации и позволяет гибко агрегировать признаки (например, «число посещений», «средний чек»), сохраняя целостность данных. E: Уникальность строк в обучающей выборке (по ID клиента) предотвращает дублирование наблюдений, которое может исказить оценку коэффициентов модели и завысить её точность.	

10	Тема 2 Решение задач статистического анализа с использованием программных пакетов.	**ПК-1**	**ИПК-1.1**	*Задания открытого типа с развернутым ответом*	10. Почему грамотное проектирование базы данных (включая нормализацию, использование внешних ключей и доменных ограничений) является необходимым условием для корректного решения задач статистического анализа в таких средах, как R или Python? Ответ: Грамотное проектирование базы данных обеспечивает целостность, согласованность и полноту исходных данных, что напрямую влияет на достоверность результатов статистического анализа. Нормализация устраняет избыточность и аномалии обновления, предотвращая дублирование информации и искажение частотных распределений. Внешние ключи гарантируют корректные связи между сущностями (например, «клиент — заказ»), что критично при объединении таблиц для формирования единой аналитической выборки. Доменные ограничения (NOT NULL, CHECK, ENUM) минимизируют количество пропущенных, ошибочных или несопоставимых значений, снижая трудозатраты на предварительную очистку данных в pandas или dplyr. Без этих мер даже самые	10. Как использование измерительных моделей (например, звёздной схемы) при проектировании базы данных упрощает и ускоряет решение задач описательной и инференциальной статистики с помощью программных пакетов, таких как statsmodels, scikit-learn или ggplot2? Ответ: Звёздная схема, состоящая из центральной таблицы фактов и окружающих её таблиц измерений (dimensions), специально оптимизирована для аналитических запросов. Она позволяет быстро агрегировать количественные показатели (например, выручка, количество транзакций) по нужным категориям (регион, время, продукт) без сложных многоуровневых JOIN-операций, характерных для полностью нормализованных OLTP-систем. Это упрощает экспорт «плоской» структурированной выборки, идеально подходящей для загрузки в DataFrame (Python) или data.frame (R). Благодаря чёткому разделению метрик и измерений, аналитик легко формирует признаки для описательной статистики (средние, дисперсии, распределения) и инференциальных моделей (регрессия, ANOVA). Кроме того, согласованные измерения (например,	

					продвинутые статистические пакеты будут работать с «грязными» или неполными данными, что приведёт к ошибочным выводам и невалидным моделям.	единый календарь) обеспечивают корректность временных и категориальных сравнений, что особенно важно при визуализации в ggplot2 или интерпретации коэффициентов в statsmodels.	
11	Тема 1. Визуализация данных с использованием пакетов статистического анализа.	**ПК-1**	**ИПК-1.2**	*Расчетная задача*	***11.*** Вы проектируете базу данных для хранения результатов опроса 1000 респондентов. Каждый респондент указывает: • возраст (целое число от 18 до 90), • пол («М» или «Ж»), • оценку удовлетворённости (от 1 до 5). Определите минимальный объём памяти (в байтах), необходимый для хранения одной записи, если использовать следующие типы данных: • TINYINT (1 байт) для возраста и оценки, • CHAR(1) (1 байт) для пола. Сколько всего памяти (в килобайтах) потребуется для хранения всех 1000 записей без учёта служебной информации СУБД? Решение: Одна запись: 1 (возраст) + 1 (пол) + 1 (оценка) = 3 байта 1000 записей: 1000 × 3 = 3000 байт ≈ 2.93 КБ	***11.*** При разработке API для передачи данных в статистическую систему вы решили использовать JSON-формат. Каждая запись содержит поля: "id" (целое число), "score" (вещественное число с точностью до двух знаков после запятой). Пример записи: {"id": 42, "score": 87.65} Оцените среднюю длину одной записи в символах (с пробелами и знаками препинания). Сколько таких записей можно передать за один вызов API, если ограничение на размер ответа — 1024 байта, а кодировка — UTF-8 (1 символ = 1 байт)? Решение: Длина примерной записи: {"id": 42, "score": 87.65} → 28 символов = 28 байт Максимум записей: ⌊1024 / 28⌋ = 36 записей	**Высокий (5-10 минут)**

<table>
<tr>
<td>12</td>
<td>Тема 1. Визуализация данных с использованием пакетов статистического анализа.</td>
<td>ПК-1</td>
<td>ИПК-1.2</td>
<td>Расчетная задача</td>
<td>12. Для построения столбчатой диаграммы в ggplot2 (R) или seaborn (Python) требуется табличная структура данных, где каждая строка — отдельное наблюдение. У вас есть данные о продажах трёх товаров (A, B, C) в двух регионах (Север, Юг):
• Товар A: Север — 120, Юг — 80
• Товар B: Север — 90, Юг — 110
• Товар C: Север — 70, Юг — 130
Сколько строк должно быть в «длинном» (long) формате таблицы, пригодной для визуализации?
Ответ:
В длинном формате каждая комбинация «Товар–Регион» — отдельная строка. 3 товара × 2 региона = 6 наблюдений. Пример структуры:
<table>
<tr><th>товар</th><th>регион</th><th>продажи</th></tr>
<tr><td>A</td><td>Север</td><td>120</td></tr>
<tr><td>A</td><td>Юг</td><td>80</td></tr>
<tr><td>B</td><td>Север</td><td>90</td></tr>
<tr><td>B</td><td>Юг</td><td>110</td></tr>
<tr><td>C</td><td>Север</td><td>70</td></tr>
<tr><td>C</td><td>Юг</td><td>130</td></tr>
</table>
Эта структура корректно поддерживается функциями geom_bar() (ggplot2) и sns.barplot() (seaborn).</td>
<td>12. Для построения тепловой карты корреляции в Python с помощью seaborn.heatmap() используется квадратная матрица корреляций размером n × n, где n — число числовых переменных.
Если у вас есть DataFrame с 5 числовыми столбцами (x1, x2, x3, x4, x5), сколько элементов будет содержать матрица корреляций, и сколько из них уникальных (без учёта диагонали и симметрии)?
Ответ:
• Всего элементов: 25
• Уникальных недиагональных элементов: 10
Обоснование:
Матрица 5×5 → $5^2 = 25$ элементов.
Диагональ содержит 5 единиц (корреляция переменной с самой собой).
Остальные 20 элементов симметричны относительно диагонали, поэтому уникальных пар: 20 / 2 = 10.
Именно эти 10 значений определяют структуру тепловой карты, так как sns.heatmap() отображает всю матрицу, но содержательно важны только уникальные корреляции.</td>
<td>Высокий (5-10 минут)</td>
</tr>
<tr>
<td>13</td>
<td>Тема 2 Решение</td>
<td>ПК-1</td>
<td>ИПК-1.2</td>
<td>Расчетная</td>
<td>13. Требуется разработать модуль для автоматического расчёта</td>
<td>13. При анализе данных необходимо визуализировать распределение</td>
<td>Высокий (5-10 минут)</td>
</tr>
</table>

	задач статисти ческого анализа с использ ованием програм мных пакетов.			*задача*	описательных статистик (среднее, медиана, стандартное отклонение) по числовым данным из CSV-файла. Какой подход к проектированию ПО следует применить, чтобы обеспечить повторное использование кода и лёгкую интеграцию с другими модулями? Ответ: Следует использовать объектно-ориентированный подход с выделением класса StatAnalyzer, инкапсулирующего методы загрузки данных и расчёта статистик. Для обеспечения повторного использования и расширяемости применяются принципы SOLID, а взаимодействие с файловой системой выносится в отдельный сервис. Реализация может быть выполнена на Python с использованием библиотек pandas и NumPy.	признака и проверить его нормальность с помощью критерия Шапиро–Уилка. Как организовать архитектуру программного модуля, чтобы отделить логику анализа от представления результатов? Ответ: Рекомендуется применить паттерн MVC (Model–View–Controller): • Model реализует статистические вычисления (например, через SciPy); • View отвечает за визуализацию (например, через Matplotlib или Seaborn); • Controller координирует взаимодействие между ними. • Такая архитектура упрощает тестирование, поддержку и замену компонентов без изменения остальной части системы.	
14	Тема 2 Решение задач статисти ческого анализа с использ ованием	**ПК-1**	**ИПК-1.2**	*Расчет ная задача*	***14.*** Из реляционной базы данных PostgreSQL таблицы sales (поля: id, product_id, amount, sale_date) требуется извлечь среднее значение продаж (amount) за последний квартал и рассчитать стандартное отклонение. Как это реализовать с использованием Python и библиотеки pandas?	***14.*** Через REST API (например, https://api.example.com/data) доступны ежедневные данные о посещаемости сайта в формате JSON: Json [{"date": "2026-01-01", "visitors": 1200}, {"date": "2026-01-02", "visitors": 1350}, ...] Требуется загрузить данные за	**Высокий (5-10 минут)**

	программных пакетов.				Ответ: python 1.import pandas as pd 2.import psycopg2 3.from sqlalchemy import create_engine 4. *5.# Подключение к БД* engine = create_engine('postgresql://user:password@localhost/dbname') 7. *8.# SQL-запрос за последний квартал (пример для Q4 2025)* 9.query = """ 10. SELECT amount 11. FROM sales 12. WHERE sale_date >= '2025-10-01' AND sale_date <= '2025-12-31'; 13.""" 14. 15.df = pd.read_sql(query, engine) 16. 17.mean_sales = df['amount'].mean() 18.std_sales = df['amount'].std() 19. 20.print(f"Среднее: {mean_sales:.2f}, Стандартное отклонение: {std_sales:.2f}") Результат (пример): Среднее: 1245.60, Стандартное отклонение: 320.75	январь 2026 г., сохранить их в SQLite-таблицу traffic, а затем вычислить медиану и 95%-й перцентиль числа посетителей. Ответ: python 1.import requests 2.import sqlite3 3.import pandas as pd 4 *5.# Получение данных через API* 6.response = requests.get("https://api.example.com/data?month=2026-01") 7.data = response.json() 8. *9.# Сохранение в SQLite* 10conn = sqlite3.connect("analytics.db") 11.df = pd.DataFrame(data) 12.df.to_sql("traffic", conn, if_exists="replace", index=False) 13. *14.# Статистический анализ* 15.median_vis = df['visitors'].median() 16.p95_vis = df['visitors'].quantile(0.95) 17. 18.print(f"Медиана: {median_vis}, 95%-й перцентиль: {p95_vis:.1f}") Результат (пример): Медиана: 1280, 95%-й перцентиль: 1840.0	

15	Тема 2 Решение задач статистического анализа с использованием программных пакетов.	**ПК-1**	**ИПК-1.2**	*Расчетная задача*	***15.*** Необходимо разработать многократно используемый модуль на Python для расчёта корреляционной матрицы по числовым данным из CSV-файла. Как применить принципы проектирования ПО (например, инкапсуляцию и разделение ответственности), чтобы модуль легко тестировался и интегрировался в другие проекты? Решение: Применим объектно-ориентированное проектирование и выделим отдельный класс CorrelationAnalyzer, соблюдая принцип единственной ответственности: python 1.import pandas as pd 2. 3.class CorrelationAnalyzer: 4. def __init__(self, file_path: str): 5. self.data = pd.read_csv(file_path) 6. 7. def compute_correlation_matrix(self, method='pearson'): 8. numeric_data = self.data.select_dtypes(include='number') 9. return	***15.*** Требуется создать консольное приложение, которое: • Загружает данные из файла, • Выполняет базовый статистический анализ (среднее, дисперсия, гистограмма), • Сохраняет результаты в отчёт. Как спроектировать архитектуру приложения, чтобы обеспечить масштабируемость и лёгкую замену компонентов (например, вывода в файл или на экран)? Решение: Используем паттерн «Стратегия» и модульную архитектуру: python 1.# analysis.py 2. import pandas as pd 3. import matplotlib.pyplot as plt 4. 5. class DataProcessor: 6. def __init__(self, loader, reporter): 7. self.loader = loader # стратегия загрузки 8. self.reporter = reporter # стратегия отчёта 9. 10. def run(self, source): 11. df = self.loader.load(source) 12. stats = { 13. 'mean': df.mean(numeric_only=True).to_dict(), 14. 'var': df.var(numeric_only=True).to_dict()	**Высокий (5-10 минут)**

					numeric_data.corr(method=method) 10. 11.# Пример использования 12.analyzer = CorrelationAnalyzer('data.csv') 13.corr_matrix = analyzer.compute_correlation_matrix() 14. print(corr_matrix) Такой подход обеспечивает: • чёткое разделение логики загрузки данных и анализа, • возможность подмены источника данных (через внедрение зависимости), • простоту unit-тестирования метода compute_correlation_matrix.	15. } 16. self.reporter.generate(stats, df) 17 18.# loaders.py 19.class CSVLoader: 20. def load(self, path): 21. return pd.read_csv(path) 22. 23.# reporters.py 24.class FileReporter: 25. def generate(self, stats, df): 26. with open('report.txt', 'w') as f: Такая архитектура: • позволяет легко заменить формат входных данных (CSV → JSON → БД), • поддерживает разные типы отчётов (файл, email, веб-интерфейс), • соответствует принципам гибкого проектирования ПО (SOLID, loose coupling).	

Критерии оценивания диагностической работы:

Оценка	Критерии оценивания
отлично	от 90 до 100 % правильно выполненных заданий
хорошо	от 70 до 89 % правильно выполненных заданий
удовлетворительно	от 50 до 69 % правильно выполненных заданий
неудовлетворительно	менее 50 % правильно выполненных заданий