

МИНОБРНАУКИ РОССИИ
ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА МАТЕМАТИКИ И МОДЕЛИРОВАНИЯ

Рабочая программа дисциплины (модуля)
СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ

Направление и направленность (профиль)
38.03.06 Торговое дело. Международная логистика и управление поставками

Год набора на ОПОП
2024

Форма обучения
очно-заочная

Владивосток 2025

Рабочая программа дисциплины (модуля) «Статистический анализ данных» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 38.03.06 Торговое дело (утв. приказом Минобрнауки России от 12.08.2020г. №963) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

*Завалин Г.С., начальник отдела, Отдел интеллектуального анализа данных,
Georgiy.Zavalin08@vvsu.ru*

*Кочева Е.В., кандидат экономических наук, доцент, Кафедра математики и
моделирования, Kocheva.EV@vvsu.ru*

*Шумик Е.Г., кандидат экономических наук, заведующий кафедрой, Кафедра
информационной безопасности, Ekaterina.Shumik1@vvsu.ru*

Утверждена на заседании кафедры математики и моделирования от 15.05.2025 ,
протокол № 9

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Галимзянова К.Н.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1599657997
Номер транзакции	0000000000E7A62F
Владелец	Галимзянова К.Н.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Целью преподавания дисциплины «Статистический анализ данных» является углубление знаний статистики, расширение умений и навыков практического решения задач, возникающих в гуманитарных и социально-экономических науках, связанных с обработкой многомерных данных.

- ознакомить студентов с понятийным аппаратом дескриптивного статистического анализа и доверительного оценивания;

- научить студентов вычислять точечные оценки неизвестных параметров распределений и строить доверительные интервалы;

- дать студентам представление об основах корреляционного анализа; - ознакомить студентов с понятийным аппаратом и критериями проверки статистических гипотез;

- выработать у студентов навыки применения статистического пакета R в прикладном статистическом анализе.

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине	
			Код результата	Формулировка результата
38.03.06 «Торговое дело» (Б-ТД)				

В процессе освоения дисциплины решаются задачи воспитания гармонично развитой, патриотичной и социально ответственной личности на основе традиционных российских духовно-нравственных и культурно-исторических ценностей, представленные в таблице 1.2.

Таблица 1.2 – Целевые ориентиры воспитания

Воспитательные задачи	Формирование ценностей	Целевые ориентиры
Формирование гражданской позиции и патриотизма		
Формирование чувства гордости за достижения России	Патриотизм Коллективизм	Внимательность к деталям Гибкость мышления Активная жизненная позиция
Формирование духовно-нравственных ценностей		
Формирование ответственного отношения к труду	Достоинство Коллективизм Справедливость	Дисциплинированность Активная жизненная позиция
Формирование научного мировоззрения и культуры мышления		
Развитие познавательного интереса и стремления к знаниям	Достоинство Взаимопомощь и взаимоуважение	Гибкость мышления Креативное мышление

Формирование коммуникативных навыков и культуры общения		
Формирование навыков публичного выступления и презентации своих идей	Коллективизм Взаимопомощь и взаимоуважение	Умение рефлексировать Умение работать в команде и взаимопомощь

2 Место дисциплины (модуля) в структуре ОПОП

Дисциплина «Статистический анализ данных» входит в базовую часть Блока 1 Дисциплины (модули) учебного плана.

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обуче- ния	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо- емкость	Объем контактной работы (час)						СРС	Форма аттес- тации
				(З.Е.)	Всего	Аудиторная			Внеауди- торная			
						лек.	прак.	лаб.	ПА	КСР		
38.03.06 Торговое дело	ОЗФО	Б1.Б	4	2	5	2	0	0	1	2	67	3

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОЗФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОЗФО

№	Название темы	Код ре-зультата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Введение в данные	РД1, РД2, РД3	1	0	1	0	Тестовые задания, практические работы
2	Статистический анализ выборочных данных	РД1, РД2, РД3	1	0	1	2	Тестовые задания, практические работы
3	Проверка статистических гипотез	РД1, РД2, РД3	2	0	2	0	Тестовые задания, практические работы
4	Сравнение и создание групп	РД1, РД2, РД3	2	0	0	0	Тестовые задания, практические работы
5	Исследование статистических взаимосвязей	РД1, РД2, РД3	2	0	0	0	Тестовые задания, практические работы
6	Актуальность и роль визуализации данных в аналитике больших данных	РД1, РД2, РД3	2	0	0	0	Тестовые задания, практические работы

Итого по таблице		10	0	4	2	
------------------	--	----	---	---	---	--

4.2 Содержание разделов и тем дисциплины (модуля) для ОЗФО

Тема 1 Введение в данные .

Содержание темы: Случайные величины и их свойства. Основные распределения случайных величин. Типы шкал измерения признаков и основные ограничения.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция- дискуссия.

Виды самостоятельной подготовки студентов по теме: Изучение рекомендованной литературы.

Тема 2 Статистический анализ выборочных данных .

Содержание темы: Генеральная совокупность и выборка. Выборочные аналоги параметров генеральной совокупности. Применение выборочного метода в различных сферах практической деятельности. Проектирование выборочных обследований и способы формирования выборки. Распространение результатов выборочного обследования на генеральную совокупность. Выборка случайных величин. Меры центральной тенденции. Меры вариативности. Описательные статистики. Ошибки выборки. Работа с пропущенными наблюдениями.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция- дискуссия.

Виды самостоятельной подготовки студентов по теме: Изучение рекомендованной литературы.

Тема 3 Проверка статистических гипотез .

Содержание темы: Статистическая гипотеза. Статистические критерии. Алгоритм проверки статистических гипотез. Свойства критериев. Критерии согласия.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция- дискуссия.

Виды самостоятельной подготовки студентов по теме: Изучение рекомендованной литературы.

Тема 4 Сравнение и создание групп .

Содержание темы: Одновыборочные и двухвыборочные критерии. Одновыборочные критерии сравнения средних. Сравнение двух независимых выборок. Сравнение дисперсий двух независимых выборок. Сравнение распределений двух независимых выборок. Сравнение двух связанных выборок. Критерии равенства групп. Сравнение средних для k независимых выборок: параметрический случай. Сравнение средних для k независимых выборок: непараметрический случай. Сравнение средних для нескольких связанных выборок.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция- дискуссия.

Виды самостоятельной подготовки студентов по теме: Изучение рекомендованной литературы.

Тема 5 Исследование статистических взаимосвязей .

Содержание темы: Коэффициенты корреляции. Коэффициенты ранговой корреляции. Регрессионный анализ. Выявление тенденции.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция- дискуссия.

Виды самостоятельной подготовки студентов по теме: Изучение рекомендованной литературы.

Тема 6 Актуальность и роль визуализации данных в аналитике больших данных .

Содержание темы: Влияние визуализации на процесс принятия решений. Ключевые подходы к визуализации больших данных. Обзор современных инструментов визуализации данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция- дискуссия.

Виды самостоятельной подготовки студентов по теме: Изучение рекомендованной литературы.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.

- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.

- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Иванюга, Т. В. Статистика : учебно-методическое пособие / Т. В. Иванюга. — Брянск : Брянский ГАУ, 2021 — Часть 2 : Социально-экономическая статистика — 2021. — 111 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/304628> (дата обращения: 09.09.2025). — Режим доступа: для авториз. пользователей.

2. Ковалева, М. А., Анализ данных : учебное пособие / М. А. Ковалева, Р. И. Бтемирова. — Москва : Русайнс, 2021. — 62 с. — ISBN 978-5-4365-8609-0. — URL: <https://book.ru/book/942003> (дата обращения: 09.09.2025). — Текст : электронный.

7.2 Дополнительная литература

1. Под ред. Мхитаряна В.С. АНАЛИЗ ДАННЫХ. Учебник для вузов [Электронный ресурс] , 2020 - 490 - Режим доступа: <https://urait.ru/book/analiz-dannyh-450166>

2. Соловьев, В. И., Анализ данных в экономике: Теория вероятностей, прикладная статистика, обработка и анализ данных в Microsoft Excel. : учебник / В. И. Соловьев. — Москва : КноРус, 2025. — 497 с. — ISBN 978-5-406-13693-5. — URL: <https://book.ru/book/955517> (дата обращения: 09.09.2025). — Текст : электронный.

3. Статистика : учебное пособие / Шумилина Т.В.; Газизьянова Ю.Ю. — Кинель : РИО СамГАУ, 2020 .— 223 с. — URL: <https://lib.rucont.ru/efd/718774> (дата обращения: 04.08.2025)

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Электронно-библиотечная система "BOOK.ru"
2. Электронно-библиотечная система "ЛАНЬ"
3. Электронно-библиотечная система "РУКОНТ"
4. Электронно-библиотечная система издательства "Юрайт" - Режим доступа: <https://urait.ru/>

5. Open Academic Journals Index (ОАИ). Профессиональная база данных - Режим доступа: <http://oaji.net/>

6. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: <https://www.prilib.ru/>

7. Информационно-справочная система "Консультант Плюс" - Режим доступа: <http://www.consultant.ru/>

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

- Компьютеры
- Проектор

Программное обеспечение:

- Microsoft Office Professional Plus 2019 Russian
- Microsoft Windows XP Professional w/SP3 32-bit Russian

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

КАФЕДРА МАТЕМАТИКИ И МОДЕЛИРОВАНИЯ

Фонд оценочных средств
для проведения текущего контроля
и промежуточной аттестации по дисциплине (модулю)

СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ

Направление и направленность (профиль)
38.03.06 Торговое дело. Международная логистика и управление поставками

Год набора на ОПОП
2024

Форма обучения
очно-заочная

Владивосток 2025

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
38.03.06 «Торговое дело» (Б-ТД)		

Компетенция считается сформированной на данном этапе в случае, если полученные результаты обучения по дисциплине оценены положительно (диапазон критериев оценивания результатов обучения «зачтено», «удовлетворительно», «хорошо», «отлично»). В случае отсутствия положительной оценки компетенция на данном этапе считается несформированной.

2 Показатели оценивания планируемых результатов обучения

Таблица заполняется в соответствии с разделом 1 Рабочей программы дисциплины (модуля).

3 Перечень оценочных средств

Таблица 3 – Перечень оценочных средств по дисциплине (модулю)

Контролируемые планируемые результаты обучения		Контролируемые темы дисциплины	Наименование оценочного средства и представление его в ФОС	
			Текущий контроль	Промежуточная аттестация
Очно-заочная форма обучения				
РД1	Знание : теоретические аспекты статистического исследования различных социально-экономических процессов на макро- и микроуровне, на основе познания конкретных методов и методик расчета и анализа экономических показателей	1.1. Введение в данные	Тест	Опрос
		1.2. Статистический анализ выборочных данных	Тест	Опрос
		1.3. Проверка статистических гипотез	Тест	Опрос
		1.4. Сравнение и создание групп	Тест	Опрос
		1.5. Исследование статистических взаимосвязей	Тест	Опрос
		1.6. Актуальность и роль визуализации данных в аналитике больших данных	Тест	Опрос
РД2	Умение : применения широкий инструментарий статистического анализа данных при решении профессиональных задач, практической деятельности, а также при при	1.1. Введение в данные	Практическая работа	Опрос
		1.2. Статистический анализ выборочных данных	Практическая работа	Опрос
		1.3. Проверка статистических гипотез	Практическая работа	Опрос

	нятии управленческих решений	1.4. Сравнение и создание групп	Практическая работа	Опрос
		1.5. Исследование статистических взаимосвязей	Практическая работа	Опрос
		1.6. Актуальность и роль визуализации данных в аналитике больших данных	Практическая работа	Опрос
РДЗ	Навык : проведения статистических расчетов с применением компьютерной техники и современных статистических пакетов прикладных программ	1.1. Введение в данные	Практическая работа	Опрос
		1.2. Статистический анализ выборочных данных	Практическая работа	Опрос
		1.3. Проверка статистических гипотез	Практическая работа	Опрос
		1.4. Сравнение и создание групп	Практическая работа	Опрос
		1.5. Исследование статистических взаимосвязей	Практическая работа	Опрос
		1.6. Актуальность и роль визуализации данных в аналитике больших данных	Практическая работа	Опрос

4 Описание процедуры оценивания

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Вид учебной деятельности	Оценочное средство				
	Собеседование	Тест	Практические задачи	Зачет тест	Итого
Лекции	10				10
Практическое занятие	10	10	30		50
Самостоятельная работа			20		20
Промежуточная аттестация				20	20
Итого	20	10	50	20	100

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обладает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.

от 61 до 75	«зачтено» / «удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным компетенциям, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворительно»	У студента не сформированы дисциплинарные компетенции, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

5 Примерные оценочные средства

5.1 Примерные вопросы

1. Что такое генеральная совокупность и выборка? В чем их отличие?
2. Какие виды статистических распределений вы знаете? Приведите примеры и объясните, когда они применяются.
3. Как вычисляется стандартная ошибка среднего значения? Объясните формулу и её значение.
4. Что такое доверительный интервал и как он рассчитывается? Приведите пример расчета доверительного интервала для среднего значения.
5. Какие методы проверки гипотез вы знаете? Опишите алгоритм проведения t-теста для сравнения средних значений двух независимых выборок.
6. Дайте определение понятий нулевой и альтернативной гипотезы. Приведите примеры использования этих гипотез в различных контекстах.
7. Что такое уровень значимости и критическая область? Как эти понятия связаны с ошибками первого и второго рода?
8. Опишите процесс проверки статистической гипотезы с использованием критерия Стьюдента (t-критерия). Приведите пошаговый алгоритм применения этого метода.
9. Чем отличаются односторонние и двусторонние тесты гипотез? Когда следует использовать каждый из них?
10. Что такое мощность теста и как она связана с уровнем значимости и размером выборки? Приведите пример расчета мощности теста для конкретного случая.
11. Что такое корреляция и какие типы корреляционных связей существуют? Приведите примеры положительных, отрицательных и отсутствующих корреляций.
12. Опишите коэффициент корреляции Пирсона. Как интерпретируется его значение? В каких случаях этот метод неприменим?
13. Что такое регрессионный анализ и какие задачи он решает? Приведите пример линейного уравнения регрессии и объясните, как оно используется для прогнозирования.
14. Как проводится проверка значимости коэффициента корреляции? Объясните шаги процедуры и приведите пример расчетов.
15. Что такое мультиколлинеарность и как она влияет на результаты регрессионного анализа? Предложите способы устранения мультиколлинеарности.

Краткие методические указания

Для лучшей подготовки к опросу, освоения материала и систематизации знаний по дисциплине необходимо разобрать материалы лекций по теме. Во время самостоятельной проработки лекционного материала особое внимание следует уделять возникшим вопросам, непонятным определениям, формулировкам, доказательствам.

Шкала оценки

Оценка	Баллы	Описание
5	19–20	ставится, если студент полностью освоил материал
4	16–18	ставится, если студент допускает 1-2 ошибки

3	13–15	ставится, если студент излагает материал неполно и допускает неточности в определении понятий или формулировке правил, излагает материал непоследовательно и допускает ошибки
2	0–12	ставится, если студент обнаруживает незнание ответа на соответствующие вопросы, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал

5.2 Примеры заданий для выполнения практических работ

Практическое занятие

«Статистический анализ выборочных данных»

Задача 1. Имеются следующие данные о величине полученной прибыли по 25 предприятиям транспорта, тыс. руб.:

250; 187; 213; 324; 165; 187; 154; 210; 205; 185; 193; 306; 163; 195; 109; 345; 218; 302; 183; 169; 142; 159; 345; 105; 283.

Постройте интервальный ряд распределения транспортных предприятий по величине полученной прибыли, образовав шесть групп с равными интервалами.

Задача 2. Произведите группировку данных по числу работающих, распределив их на 4 группы.

Составьте: 1) рабочую таблицу; 2) сводную таблицу и подсчитайте по каждой группе: а) частоты и частоты; б) число работающих; в) число работающих в процентах к итогу; г) выпуск продукции; д) выпуск продукции в процентах к итогу; е) производительность труда.

Постройте гистограммы: а) частот; б) числа работающих в процентах к итогу; в) выпуска продукции в процентах к итогу.

Номер предприятия	Выпуск продукции, млн. руб.	Число работающих	Номер предприятия	Выпуск продукции, млн. руб.	Число работающих
1	10400	190	13	4300	86
2	9900	178	14	38500	438
3	17100	263	15	24100	278
4	34400	343	16	7300	117
5	20000	245	17	17400	220
6	40500	410	18	43400	435
7	54500	494	19	9100	120
8	5300	98	20	4000	64
9	3800	79	21	4800	99
10	10100	182	22	7900	119
11	23300	307	23	40100	405
12	13200	224	24	32400	336

Практическое занятие «Статистические гипотезы»

Задача 1

Условие: В компании утверждают, что средний возраст сотрудников составляет 35 лет. Проведите тест на уровне значимости 0.05, если в выборке из 30 сотрудников средний возраст составил 37 лет, а стандартное отклонение равно 5 лет.

Задача 2

Условие: Исследуется, влияет ли новый препарат на снижение артериального давления. В контрольной группе среднее давление составило 120 мм рт. ст., а в экспериментальной — 115 мм рт. ст. При этом стандартные отклонения составляют 10 и 8 соответственно, а размер выборок равен 25.

Практическое занятие

«Сравнение и создание групп»

Задача 1: Сравнение средних

В исследовании сравниваются два метода обучения. Группа А (30 студентов) использовала метод 1 и показала средний балл 75 с дисперсией 16. Группа В (30 студентов) использовала метод 2 и показала средний балл 82 с дисперсией 25. Проведите тест на равенство средних с уровнем значимости 0.05.

Задача 2: Сравнение пропорций

В опросе о предпочтениях в еде 200 человек выбрали вегетарианскую диету, а 300 человек — мясную. Проверить, есть ли статистически значимая разница в пропорциях между двумя группами, используя уровень значимости 0.01.

Практическое занятие

«Исследование статистических взаимосвязей»

Задача 1: Корреляционный анализ

Данные: Влияние расходов на рекламу на объем продаж.

Месяц	Расходы на рекламу (тыс. руб)	Объем продаж (тыс. руб)
Январь	50	500
Февраль	70	600
Март	60	550
Апрель	80	700
Май	90	750
Июнь	100	800

Задача: Рассчитайте коэффициент корреляции Пирсона между расходами на рекламу и объемом продаж. Есть ли статистически значимая связь?

Задача 2: Регрессионный анализ

Данные: Влияние цены на товар и количества проданных единиц.

Цена (руб)	Количество проданных единиц
100	200
150	150
200	100
250	50
300	20

Задача: Постройте линейную регрессионную модель для предсказания количества проданных единиц на основе цены товара.

Практическое занятие

«Современные инструменты визуализации данных с использованием BI-систем»

1. Подготовка набора данных:

Используйте следующий встроенный набор данных о продажах и расходах компании за 12 месяцев:

Месяц	Продажи (тыс. руб)	Расходы (тыс. руб)	Количество клиентов
Январь	500	300	50
Февраль	600	350	60
Март	550	320	55
Апрель	700	400	70
Май	800	450	80
Июнь	750	420	75
Июль	900	500	90
Август	950	550	95
Сентябрь	800	480	85
Октябрь	850	490	88
Ноябрь	1000	600	100
Декабрь	1100	650	110

1. Анализ данных:

Проанализируйте данные, чтобы выявить ключевые тенденции:

- Как изменяются продажи и расходы в течение года?
- Какое количество клиентов связано с объемом продаж?

1. Создание визуализаций:

Создайте следующие визуализации:

- Линейный график для отображения изменения продаж и расходов по месяцам.
- Столбчатая диаграмма для сравнения количества клиентов с объемом продаж.
- Круговая диаграмма для отображения доли каждого месяца в общих продажах за год.

1. Создание дашборда:

- Объедините созданные визуализации на одном дашборде.
- Добавьте интерактивные элементы, если это возможно (например, фильтры по месяцам).

1. Подготовка отчета:

Напишите краткий отчет (1-2 страницы), в котором:

- Опишите цели анализа.
- Приведите основные выводы на основе визуализаций.
- Объясните, как данные могут помочь в принятии бизнес-решений.

1. Презентация результатов:

- Подготовьте презентацию (5-10 слайдов) для представления ваших результатов коллегам или преподавателю.
- Включите в презентацию ключевые визуализации и выводы.

Краткие методические указания

При подготовке к практической работе особое внимание необходимо обратить на уяснение практического приложения рассматриваемых теоретических вопросов по соответствующей теме. В процессе этой работы студент должен стремиться понять и запомнить основные положения рассматриваемого теоретического материала, примеры, поясняющие его, а также разобраться в методах решения задач.

Шкала оценки

Оценка	Баллы	Описание
5	39–50	Задания выполнены полностью и правильно
4	26–38	Задания выполнены полностью, с несущественными ошибками, но подход к решению, идея решения, метод правильны
3	13–25	Задания выполнены полностью, с существенными ошибками, но подход к решению, идея решения, метод правильны
2	9–12	Задания не выполнены или выполнены неправильно

5.3 Примеры тестовых заданий

РАЗДЕЛ I. ЗАДАНИЯ УРОВНЕЙ «ЗНАТЬ» И «ЗНАТЬ И УМЕТЬ»

Тема 1. Введение в данные

Задания уровня «знать»

Задание закрытой формы с единственным выбором

1. Под термином «данные» в статистике понимается

- 1) только числовая информация
- 2) любая упорядоченная информация о наблюдаемых объектах
- 3) результаты экспериментов
- 4) любые тексты, изображения и звуки

2. К типу числовых (количественных) данных относится
- 1) возраст человека
 - 2) название города
 - 3) цвет автомобиля
 - 4) категория товара
3. Категориальные переменные характеризуются тем, что
- 1) выражаются только числами
 - 2) описывают количественные характеристики
 - 3) принимают значения в виде категорий или классов
 - 4) изменяются по непрерывной шкале
4. В терминах статистики переменная представляет собой
- 1) единицу измерения данных
 - 2) характеристику объекта, принимающую разные значения
 - 3) конкретное значение данных
 - 4) ошибку в измерениях
5. Целью первичного анализа данных является
- 1) построение модели прогнозирования
 - 2) очистка и интерпретация данных
 - 3) применение статистических тестов
 - 4) визуализация больших массивов информации
6. Термин «выборка» в статистическом анализе данных означает
- 1) совокупность всех данных в исследовании
 - 2) часть данных, отобранных для анализа
 - 3) процесс измерения данных
 - 4) ошибка при анализе данных
7. Генеральная совокупность в статистике представляет собой
- 1) все возможные объекты исследования
 - 2) данные, отобранные случайным образом
 - 3) ошибки в эксперименте
 - 4) признаки исследуемого объекта
8. Шкала измерений, которая позволяет только ранжировать объекты
- 1) номинальная
 - 2) порядковая
 - 3) интервальная
 - 4) отношений
9. В данном формате чаще всего хранятся большие объемы данных
- 1) JSON
 - 2) XML
 - 3) Excel
 - 4) SQL
10. Статистическая характеристика, которая определяет среднее значение данных
- 1) квартиль
 - 2) размах вариации
 - 3) среднее арифметическое
 - 4) дисперсия
11. Выбросами в данных при проведении статистического анализа называют
- 1) ошибки в измерениях
 - 2) значения, значительно отличающиеся от остальных
 - 3) средние значения
 - 4) данные, отсутствующие в выборке
12. Значение медианы в выборке показывает
- 1) среднее значение

- 2) самое частое значение
 3) центральное значение, делящее данные пополам
 4) разницу между максимальным и минимальным значением

13. Популярный метод, который применяется для нормализации данных

- 1) логарифмирование
 2) масштабирование
 3) усреднение

4) кодирование категорий

14. Номинальная шкала измерения характеризуется

- 1) наличием нуля и равными интервалами
 2) порядком значений
 3) только возможностью классификации

4) возможностью проведения математических операций

15. Метрика, которая используется для измерения разброса данных

- 1) среднее
 2) дисперсия
 3) медиана

4) корреляция

16. «Репрезентативность выборки» в статистическом анализе данных означает

- 1) наличие ошибок в выборке
 2) соответствие выборки характеристикам генеральной совокупности
 3) случайное распределение данных

4) малый объем данных

17. Стандартное отклонение в статистике определяется

- 1) средним значением данных
 2) измерением отклонения данных от среднего
 3) процентным соотношением выборки

4) типом статистического анализа

18. Для оценки взаимосвязи между двумя переменными используется метод

- 1) корреляционного анализа
 2) кластерного анализа
 3) дисперсионного анализа

4) описательных статистик

19. Данные в контексте аналитики представляют собой

- 1) любую информацию, хранящуюся в цифровом виде
 2) числовые значения, собранные в таблице
 3) структурированную или неструктурированную информацию, используемую для анализа

4) только текстовые и числовые значения

20. К категориальным (качественным) данным относится

- 1) возраст человека
 2) температура воздуха
 3) марка автомобиля

4) рост человека

Задания закрытой формы с **множественным выбором**

21. К основным типам данных в статистике относят

- 1) числовые
 2) категориальные
 3) дискретные

4) логарифмические

22. К категориальным переменным относятся

- 1) пол человека
 2) средняя температура

- 3) цвет глаз
- 4) рост человека
23. Для визуализации данных в статистике используются методы
- 1) гистограмма
- 2) таблица
- 3) дисперсионного анализа
- 4) диаграмма рассеяния
24. К числовым типам данным относятся
- 1) дискретные данные
- 2) непрерывные данные
- 3) категориальные данные
- 4) номинальные данные
25. Форматы файлов, которые чаще всего используются для хранения табличных данных
- 1) CSV
- 2) JSON
- 3) TXT
- 4) XLSX
26. Для оценки качества данных относятся следующие характеристики
- 1) полнота данных
- 2) точность данных
- 3) дублирование данных
- 4) актуальность данных
27. Операции, которые относятся к этапу предобработки данных
- 1) очистка данных
- 2) визуализация данных
- 3) нормализация данных
- 4) обучение модели
28. Шкалы, которые используются для измерения данных
- 1) номинальная
- 2) порядковая
- 3) интервальная
- 4) дихотомическая
29. Задачи, которые можно решать с помощью анализа данных
- 1) поиск закономерностей
- 2) прогнозирование тенденций
- 3) изменение исходных данных
- 4) оптимизация бизнес-процессов
30. Источники данных, используемые в аналитике больших данных
- 1) базы данных
- 2) социальные сети
- 3) сенсоры и датчики
- 4) газеты и журналы

Задания на установление соответствия

31. Установите соответствие между шкалами измерения и их характеристиками:

- 1) номинальная
- 2) порядковая
- 3) интервальная
- 4) отношений
- А) отражает порядок значений
- В) позволяет измерять разницу

- C) не имеет естественного порядка
- D) имеет нулевую точку отсчета

32. Установите соответствие между метриками описательной статистики и их определениями:

- 1) дисперсия
- 2) медиана
- 3) мода
- A) среднее квадратичное отклонение
- B) наиболее часто встречающееся значение
- C) центральное значение выборки

33. Установите соответствие между типами переменных и примерами

- 1) дискретная переменная
- 2) непрерывная переменная
- 3) категориальная переменная
- A) цвет автомобиля
- B) температура воздуха
- C) количество учеников в классе

34. Установите соответствие между методами обработки данных и их определениями

- 1) кодирование
- 2) нормализация
- 3) дискретизация
- 4) импутация
- A) заполнение пропущенных значений
- B) приведение значений к единому масштабу
- C) разбиение непрерывных данных на интервалы
- D) преобразование категориальных данных в числовые

35. Установите соответствие между этапами работы с данными и их содержанием

- 1. определение цели анализа
- 2. сбор данных
- 3. очистка данных
- 4. анализ данных
- 5. интерпретация результатов
- A) получение информации из различных источников
- B) формулировка вопросов исследования
- C) удаление выбросов, обработка пропущенных значений
- D) выводы и рекомендации на основе анализа
- E) вычисление статистических показателей, построение моделей

Задания на **установление правильной последовательности**

36. Расположите этапы анализа данных в правильной последовательности:

- 1. сбор данных
- 2. анализ данных
- 3. очистка данных
- 4. визуализация

37. Расположите этапы предобработки данных в правильной последовательности

1. заполнение пропущенных значений
 2. удаление дубликатов
 3. преобразование форматов данных
 - 4 кодирование категориальных переменных
38. Установите последовательность этапов работы с базами данных:
1. подключение к базе данных
 2. выполнение SQL-запросов
 3. извлечение данных
 4. обработка и анализ данных
39. Упорядочите этапы работы с визуализацией данных:
1. определение цели визуализации
 2. выбор типа графика
 3. построение диаграммы
 - 4 анализ и интерпретация результатов
40. Расположите этапы подготовки данных для машинного обучения в правильном порядке:
1. сбор данных
 2. нормализация или стандартизация данных
 3. разделение данных на обучающую и тестовую выборки
 4. очистка и обработка данных
- Задание открытой формы**
41. Основной тип данных, который используется для количественного измерения, называется_____.
42. Статистический показатель, который описывает среднее отклонение от среднего значения, называется_____.
43. Метод, который используется для представления данных в виде графиков и диаграмм, называется_____.
44. Если данные изменяются на основе времени, то они являются_____.
45. Ошибка, которая возникает из-за неправильного сбора или обработки данных, называется_____.

Тема 1. Введение в данные

Задания уровня «знать и уметь»

Задание закрытой формы с единственным выбором

46. С целью сбора данных о посещаемости и проведения анализа поведения пользователей на сайте интернет-магазина проводится с использованием
- 1) Excel
 - 2) Google Analytics
 - 3) PowerPoint
 - 4) Tableau
47. В компании U необходимо предсказать спрос на такси в разные часы дня, для это лучше всего воспользоваться
- 1) текстовыми отзывами пользователей
 - 2) историческими данными о заказах
 - 3) логотипом компании
 - 4) электронными письмами сотрудников
48. Перед проведением анализа данных о продажах в Excel необходимо
- 1) удалить дубликаты
 - 2) применить сводные таблицы
 - 3) построить диаграмму
 - 4) изменить шрифт в таблице

49. Наиболее подходящий инструмент для анализа предпочтений пользователей для персонализированных рекомендаций

- 1) Power BI
- 2) Photoshop
- 3) Microsoft Word
- 4) CorelDraw

50. Для анализа показателей эффективности рекламных кампаний компании будут полезны данные

- 1) Внутренние отчеты сотрудников
- 2) История покупок пользователей
- 3) Фотографии товаров
- 4) Дизайн сайта

51. Для визуализации данных о продажах компании в динамике за 12 месяцев отчетного года лучше всего подойдет

- 1) Гистограмма
- 2) Круговая диаграмма
- 3) Линейный график
- 4) Диаграмма рассеяния

52. Для анализа спроса на напитки компании S в разных регионах необходимы данные

- 1) расположение магазинов и количество заказов по категориям
- 2) переписку сотрудников в мессенджерах
- 3) список поставщиков кофе
- 4) историю изменений меню

53. Для расчета среднего чека клиента в интернет-магазине используется формула

- 1) СУММ()
- 2) СРЗНАЧ()
- 3) МАКС()
- 4) ЕСЛИ()

54. С целью прогнозирования количественных данных используют

- 1) описательный анализ
- 2) диагностический анализ
- 3) предсказательный анализ
- 4) текстовый анализ

55. С целью построения аналитических интерактивных отчетов используется

- 1) Microsoft Word
- 2) Power BI
- 3) Google Docs
- 4) Notepad

Задания закрытой формы с **множественным выбором**

56. Источники данных, которые подходят компании общественного питания для анализа данных с целью оптимизации меню.

- 1) история заказов клиентов
- 2) отзывы пользователей в социальных сетях
- 3) финансовая отчетность конкурентов
- 4) списки ингредиентов в рецептах

57. Инструмент для построения дашбордов ключевых метрик бизнеса

- 1) Tableau
- 2) Power BI
- 3) SQL
- 4) Google Slides

58. С целью анализа предпочтений клиентов для персонализированных предложений используются методы

- 1) кластеризация пользователей
- 2) анализ временных рядов
- 3) определение среднего возраста сотрудников
- 4) разработка логотипа

59. При анализе интернет-рекламы используются данные

- 1) демографические характеристики пользователей
- 2) история просмотров сайтов
- 3) генетическая информация клиентов
- 4) данные о погоде

60. К основным этапам предварительного анализа данных относятся

- 1) проверка структуры и типов данных*
- 2) формулирование научной гипотезы
- 3) визуализация распределений переменных*
- 4) обработка пропущенных значений*

Задания на установление **соответствия**

61. Установите соответствие между компанией и видом данных, которые она активно использует:

- 1) Amazon
- 2) Netflix
- 3) Google
- 4) Uber
- A) данные о поисковых запросах пользователей
- B) данные о покупках и предпочтениях клиентов
- C) данные о просмотрах фильмов и рейтингах
- D) данные о геолокации и заказах такси

62. Установите соответствие между видом данных и примером их использования:

- 1) транзакционные данные
- 2) геолокационные данные
- 3) временные ряды
- 4) текстовые данные
- A) анализ задержек рейсов в авиакомпаниях
- B) определение популярных маршрутов в Google Maps
- C) анализ покупок клиентов в интернет-магазине
- D) анализ отзывов пользователей на сайтах

63. Установите соответствие между термином и его определением:

- 1) Big Data
- 2) Data Science
- 3) Machine Learning
- 4) SQL
- A) программное обеспечение для обработки данных
- B) область, включающая анализ и обработку данных
- C) раздел ии, связанный с обучением на данных
- D) язык запросов для работы с базами данных

64. Установите соответствие между инструментами и их основным назначением:

- 1) Tableau
- 2) Python
- 3) MySQL

4) Excel

A) визуализация данных

B) обработка и анализ больших данных

C) управление реляционными базами данных

D) работа с таблицами и базовый анализ

65. Установите соответствие между этапами работы с данными и их описанием:

1) сбор данных

2) обработка данных

3) анализ данных

4) визуализация данных

A) очистка и преобразование данных перед анализом

B) выгрузка информации из разных источников

C) поиск закономерностей и трендов

D) представление информации в виде графиков

Задания на **установление правильной последовательности**

66. Установите правильную последовательность шагов при очистке данных в Excel перед анализом:

1. применение фильтров и удаление выбросов
2. проверка данных на дубликаты
3. удаление пропущенных значений или их замена
4. приведение данных к единому формату (например, даты, числа)

67. Расположите шаги анализа данных при прогнозировании спроса в интернет-магазине:

1. сбор данных о продажах за предыдущие периоды
2. очистка данных от аномалий и ошибок
3. применение статистических методов прогнозирования
4. визуализация результатов и подготовка отчета

68. Установите последовательность действий при проведении анализа клиентской базы в CRM-системе:

1. сегментация клиентов по ключевым характеристикам
2. выгрузка данных из crm-системы
3. очистка данных и обработка пропусков
4. интерпретация результатов и формирование рекомендаций

69. Установите правильную последовательность шагов при создании автоматизированного отчета в Power BI:

1. разработка визуализаций и графиков
2. очистка и трансформация данных в power query
3. подключение источников данных
4. настройка фильтров и параметров для удобного анализа

70. Установите правильную последовательность действий маркетолога для анализа продаж продукции

1. сделать выводы о тенденциях продаж
2. визуализировать данные с помощью графиков и диаграмм
3. собрать данные о продажах за последние кварталы
4. проанализировать собранные данные на предмет закономерностей и тенденций

Задание открытой формы

71. Студент, способный правильно классифицировать данные, должен уметь определить, что данные о цвете глаз являются _____ типом данных.

72. Умение анализировать данные предполагает способность _____ данные для получения значимой информации.

73. Студент, который понимает типы данных, должен знать, что данные о росте людей являются _____ типом данных.

74. Цель статистического анализа данных _____ значимую информацию из сырых данных.

75. Умение работать с текстовыми данными предполагает способность анализировать и обрабатывать _____ данные.

Тема 2. Статистический анализ выборочных данных

Задания уровня «знать»

Задание закрытой формы с единственным выбором

76. Выборкой в статистике называется
1) все объекты, изучаемые в рамках исследования
2) часть генеральной совокупности, отобранная для анализа
3) любая совокупность данных
4) данные, собранные по всему населению

77. Метод выбора элементов выборки, при котором каждый элемент генеральной совокупности имеет равную вероятность быть выбранным
1) случайная выборка
2) кластерная выборка
3) квотная выборка
4) целевая выборка

78. Выборочное среднее как правило обозначается как
1) μ
2) σ
3) $\bar{x}$
4) s

79. Дисперсия выборки показывает
1) размах данных
2) среднее значение выборки
3) меру разброса значений относительно среднего
4) среднюю ошибку выборочного среднего

80. Стандартное отклонение выборки вычисляется как
1) квадратный корень из дисперсии
2) среднее значение выборки
3) разность максимального и минимального значений
4) сумма всех значений выборки

81. Для применения центральной предельной теоремы достаточным считается объем выборки
1) 10
2) 20
3) 30
4) 50

82. Метод выборки, который предполагает разбиение генеральной совокупности на однородные группы и случайный отбор из каждой групп
1) простая случайная выборка
2) стратифицированная выборка
3) кластерная выборка
4) квотная выборка

83. Ошибка выборки в статистике означает
- 1) отклонение выборочных характеристик от параметров генеральной совокупности
 - 2) ошибка, возникающая из-за неправильно собранных данных
 - 3) ошибка при вычислении стандартного отклонения
 - 4) разность между выборочной дисперсией и выборочным средним

84. Показатель, который оценивает степень отклонения выборочных значений от среднего

- 1) размах
- 2) дисперсия
- 3) мода
- 4) коэффициент корреляции

85. Мера, показывающая, насколько сильно изменяются значения выборки
- 1) среднее арифметическое
 - 2) медиана
 - 3) стандартное отклонение
 - 4) коэффициент вариации

86. Если в выборке распределение значений асимметрично вправо, это означает, что
- 1) среднее значение больше медианы
 - 2) среднее значение меньше медианы
 - 3) среднее и медиана равны
 - 4) выборка имеет нормальное распределение

87. Самый чувствительный параметр выборки
- 1) медиана
 - 2) мода
 - 3) среднее
 - 4) первый квартиль

88. Если коэффициент вариации менее 10%, то
- 1) данные имеют высокий разброс
 - 2) данные имеют низкий разброс
 - 3) данные не подчиняются нормальному распределению
 - 4) выборка является несбалансированной

89. Выборка, при которой каждый элемент генеральной совокупности имеет одинаковую вероятность попасть в исследование
- 1) стратифицированная
 - 2) простая случайная
 - 3) кластерная
 - 4) квотная

90. Если выборочная дисперсия равна 25, то стандартное отклонение равно:
- 1) 5
 - 2) 12.5
 - 3) 25
 - 4) 625

91. Метод, при котором исследуемая совокупность разбивается на группы, а затем отбираются
- 1) простые целые группы случайная выборка
 - 2) стратифицированная выборка
 - 3) кластерная выборка
 - 4) квотная выборка

92. Коэффициент корреляции Пирсона используется
- 1) для оценки средней величины выборки
 - 2) для измерения линейной зависимости между переменными

- 3) для вычисления медианы
- 4) для оценки ошибки выборки
93. Доверительный интервал в статистике представляет собой
- 1) границы, в которых с заданной вероятностью находится истинное значение параметра
- 2) среднее значение выборки
- 3) разность между максимальным и минимальным значением
- 4) оценка медианы выборки
94. При увеличении объема выборки доверительный интервал
- 1) уменьшается
- 2) увеличивается
- 3) не изменяется
- 4) может как увеличиваться, так и уменьшаться
95. Если уровень значимости теста составляет 0.05, это означает, что вероятность ошибки первого рода равна:
- 1) 0.01
- 2) 0.05
- 3) 0.95
- 4) 5.0

Задания закрытой формы с множественным выбором

96. Выборки, в которых отбор единиц совокупности является случайным
- 1) простая случайная
- 2) стратифицированная
- 3) квотная
- 4) кластерная
97. Эти характеристики являются мерами центральной тенденции
- 1) среднее арифметическое
- 2) дисперсия
- 3) медиана
- 4) мода
98. Вероятностными видами выборки среди перечисленных являются
- 1) простая случайная
- 2) стратифицированная
- 3) квотная
- 4) кластерная
99. Для нормального распределения верны следующие утверждения
- 1) среднее, медиана и мода совпадают
- 2) гистограмма распределения имеет форму колокола
- 3) коэффициент асимметрии всегда больше нуля
- 4) 68% наблюдений лежат в пределах одного стандартного отклонения от среднего
100. На точность выборочных оценок оказывают влияние
- 1) размер выборки
- 2) метод отбора выборки
- 3) наличие выбросов в данных
- 4) количество переменных в выборке
101. Эти показатели являются мерами разброса данных
- 1) размах
- 2) дисперсия
- 3) среднее арифметическое
- 4) стандартное отклонение
102. Для проверки статистических гипотез используются методы
- 1) критерий Стьюдента (t-тест)

- 2) дисперсионный анализ (ANOVA)
- 3) корреляционный анализ
- 4) хи-квадрат тест
103. Параметры, описывающие форму распределения
- 1) коэффициент асимметрии
- 2) коэффициент эксцесса
- 3) среднеквадратическое отклонение
- 4) дисперсия
104. Для доверительного интервала верны следующие утверждения
- 1) увеличение размера выборки уменьшает ширину доверительного интервала
- 2) более высокий уровень доверия (например, 99% вместо 95%) приводит к расширению интервала
- 3) доверительный интервал всегда симметричен
- 4) если интервал не включает 0, гипотеза о равенстве средних отвергается
105. К систематическим ошибкам в выборке могут приводить факторы
- 1) смещение при отборе участников
- 2) ошибки в измерениях
- 3) случайные колебания выборочных данных
- 4) неполнота данных
106. Для выборочного среднего верны следующие утверждения
- 1) оно является несмещенной оценкой генерального среднего
- 2) оно всегда равно генеральному среднему
- 3) оно подчиняется нормальному распределению при больших объемах выборки
- 4) оно менее устойчиво к выбросам, чем медиана
- Задания на установление **соответствия**
107. Установить соответствие между статистическими показателями и их определениями:
1. среднее арифметическое
 2. дисперсия
 3. медиана
 4. мода
 - A) значение, которое встречается чаще всего
 - B) средний квадрат отклонений от среднего
 - C) центральное значение упорядоченного ряда
 - D) сумма значений, деленная на их количество
108. Установить соответствие между видами выборок и их особенностями
1. простая случайная
 2. стратифицированная
 3. кластерная
 4. квотная
 - A) деление на группы по значимым признакам
 - B) отбор всей группы целиком
 - C) каждый элемент имеет равные шансы быть выбранным
 - D) формируется по заранее установленным критериям
109. Установить соответствие между обозначениями статистических величин и их названиями
- 1.
 2. S^2
 - 3.
 4. N
 - A) дисперсия выборки

- В) среднее выборки
- С) среднеквадратическое отклонение генеральной совокупности
- Д) объем генеральной совокупности

110. Установите соответствие между видами ошибок и их описанием:

- 1. ошибка первого рода
- 2. ошибка второго рода
- 3. систематическая ошибка
- 4. выборочная ошибка

- А) отклонение оценок из-за особенностей метода сбора данных
- В) принятие ложной гипотезы при истинной нулевой гипотезе
- С) принятие нулевой гипотезы при ложной альтернативной
- Д) отклонения, вызванные конечным размером выборки

Задания на установление правильной последовательности

111. Расположите этапы статистического анализа выборочных данных в правильном порядке:

- | | | |
|----|----------------------|-------------|
| 1. | определение | гипотезы |
| 2. | сбор | данных |
| 3. | анализ | данных |
| 4. | интерпретация | результатов |
| 5. | формирование выборки | |

112. Расположите шаги расчета стандартного отклонения в правильном порядке:

- 1. найти среднее выборки
- 2. вычислить разности между каждым значением и средним
- 3. возвести разности в квадрат
- 4. извлечь квадратный корень
- 5. найти среднее квадратов разностей

Задание открытой формы

113. Средняя ошибка выборочного среднего уменьшается при увеличении ____.

114. Выборка, в которой каждый элемент попадает в исследование по заранее установленной квоте, называется ____.

115. Центральная предельная теорема утверждает, что при увеличении объема выборки распределение выборочного среднего стремится к ____.

Тема 2. Статистический анализ выборочных данных

Задания уровня «знать и уметь»

Задание закрытой формы с единственным выбором

116. Мода выборки: 4, 7, 5, 7, 6 составляет

- | | |
|------|---|
| 1) | 4 |
| 2) | 5 |
| 3) | 6 |
| 4) 7 | |

117. Медиана выборки: 12, 18, 15, 10, 14 составляет

- | | |
|-------|----|
| 1) | 10 |
| 2) | 14 |
| 3) | 15 |
| 4) 18 | |

118. Дисперсия выборки: 3, 5, 7, 9, 11 составляет

- | | |
|----|----|
| 1) | 8 |
| 2) | 10 |

- 3) 12
- 4) 15
119. Среднее квадратическое отклонение выборки: 2, 4, 6, 8, 10 составляет
- 1) 2.83
- 2) 3.16
- 3) 4.00
- 4) 5.24
120. Выборочное среднее выборки: 20, 25, 30, 35, 40 составляет
- 1) 25
- 2) 30
- 3) 32
- 4) 35

Задания закрытой формы с множественным выбором

121. Средний рост студентов в выборке из 100 человек составил 175 см, а стандартное отклонение – 6 см, тогда 95%-й доверительный интервал для истинного среднего роста студентов (квантиль нормального распределения $z = 1.96$) нижняя и верхняя граница интервала соответствует

- 1) нижняя граница: 173,82
- 2) верхняя граница: 220,15
- 3) нижняя граница: 99,99
- 4) верхняя граница: 176,18

122. По выборке: 8, 12, 15, 20, 25, 30, 35, 40, 50 первый (Q1) и третий (Q3) квартиль составляют

- 1) Q1 = 18,3
- 2) Q3 = 37,5
- 3) Q1 = 13,5
- 4) Q3 = 40,2

123. Дисперсия и среднее квадратическое отклонение для выборки: 4, 6, 8, 10, 12 составляют

- 1) дисперсия: 8
- 2) среднее квадратическое отклонение: 2,8
- 3) дисперсия: 5
- 4) среднее квадратическое отклонение: 2,5

124. Мода и медиана выборки: 3, 3, 5, 7, 8, 8, 8, 9, 10 соответствуют значениям

- 1) мода: 8
- 2) медиана: 9
- 3) мода: 5
- 4) медиана: 8

125. В выборке из 50 наблюдений средний доход составил 45 000 руб., а стандартное отклонение – 5 000 руб. По данным средняя ошибка выборки и предельная ошибка при 95%-м доверительном уровне ($z = 1.96$) соответствуют значениям

- 1) средняя ошибка выборки: 800,8
- 2) средняя ошибка выборки: 707,1
- 3) предельная ошибка: 1381,9
- 4) предельная ошибка: 1000,2

126. В выборке из 64 человек средний вес составил 70 кг, а стандартное отклонение – 8 кг. Тогда нижняя и верхняя границы 99%-й доверительного интервала для среднего веса (квантиль нормального распределения $z = 2.576$) соответствуют значениям

- 1) нижняя граница: 100,17
- 2) верхняя граница: 72,58

3) нижняя граница: 67,42
 4) верхняя граница: 150,12

127. Значения первого (Q1) и третьего (Q3) квартилей по выборке 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 составляют соответственно

1) Q1 = 10,2
 2) Q3 = 42,5
 3) Q1 = 12,5
 4) Q3 = 30,1

128. Значение дисперсии и стандартного отклонения по выборке: 2, 4, 6, 8, 10, 12 составляют

1) дисперсия: 11,67
 2) среднее квадратическое отклонение: 3,42
 3) дисперсия: 16,21
 4) среднее квадратическое отклонение: 5,25

129. Значения моды и медианы для выборки: 1, 3, 3, 3, 5, 6, 7, 8, 9, 9 составляют соответственно

1) мода: 3
 2) медиана: 3
 3) мода: 6
 4) медиана: 5,5

130. В выборке из 36 человек средний балл по экзамену составил 80, а стандартное отклонение – 12. Средняя ошибка выборки и предельная ошибка при 95%-м доверительном уровне ($z = 1.96$) составят

1) средняя ошибка выборки: 2
 2) средняя ошибка выборки: 17
 3) предельная ошибка: 3,92
 4) предельная ошибка: 5,74

Задания на установление соответствия

131. Установите соответствие между статистической характеристикой и ее значением по данным интервального ряда.

Интервал	Частота
10 – 20	3
20 – 30	5
30 – 40	7
40 – 50	4
50 – 60	1

1) мода
 2) медиана
 3) выборочная средняя
 А) 32,86
 В) 34,00
 С) 32,50

132. Установите соответствие между статистической характеристикой и ее значением по данным интервального ряда

Интервал	Частота
5 – 15	4
15 – 25	6
25 – 35	10
35 – 45	5

- 1) дисперсия
- 2) среднее квадратическое отклонение
- 3) мода
- A) 29,44
- B) 12,5
- C) 3,54

133. Установите соответствие между статистической характеристикой и её значением по данным интервального ряда

Интервал	Частота
1 – 10	2
10 – 20	8
20 – 30	12
30 – 40	6

- 1) медиана
- 2) среднее значение
- 3) размах вариации
- 4) мода
- A) 21,00
- B) 39,00
- C) 17,50
- D) 24,00

Задания на установление правильной последовательности

134. Расположите этапы вычисления доверительного интервала в правильной последовательности

1. вычислить стандартную ошибку среднего
2. определить квантиль нормального распределения (z) в зависимости от доверительного уровня
3. найти границы интервала
4. найти выборочное среднее

135. Расположите этапы вычисления первого и третьего квартиля в правильном порядке:

1. определить позиции квартилей в отсортированной выборке
2. вычислить среднее двух соседних значений, если позиция нецелая
3. отсортировать выборку по возрастанию
4. найти значения квартилей, соответствующие вычисленным позициям

136. Расположите этапы вычисления дисперсии и стандартного отклонения в правильном порядке:

1. возвести отклонения в квадрат и найти их сумму
2. найти выборочное среднее
3. разделить сумму квадратов отклонений на количество элементов (или на $n-1$ для выборочной дисперсии)
4. вычесть среднее из каждого значения выборки
5. найти квадратный корень из дисперсии для получения стандартного отклонения

Задание открытой формы

137. Если в выборке из 49 наблюдений, а среднее значение равно 25 и стандартное отклонение – 7, то верхняя граница 95%-го доверительного интервала (квантиль нормального распределения $z = 1.96$) будет равна (ответ округлите до целого числа)_____.

138. Дана выборка: 8, 10, 12, 14, 16. Определите значение стандартного отклонения (округлите до двух знаков после запятой) _____.

139. В выборке из 100 элементов средний доход составил 50 000 руб, а стандартное отклонение – 8 000 руб. Найдите выборочную ошибку среднего (округлите до целого числа) _____.

140. Если дан ряд чисел: 7, 12, 15, 18, 22, 26, 30, то медиана равна _____.

Тема 3. Проверка статистических гипотез

Задания уровня «знать»

Задание закрытой формы с **единственным выбором**

141 Для проверки гипотезы о среднем используется критерий

- 1) критерий хи-квадрат
- 2) t-критерий Стьюдента
- 3) критерий Фишера
- 4) критерий Пирсона

142. Нулевой гипотезой называется гипотеза

- 1) которую мы пытаемся доказать
- 2) которую мы пытаемся опровергнуть
- 3) которая является альтернативной
- 4) которая всегда верна

143. Значение р-уровня в тестировании гипотез означает

- 1) вероятность того, что нулевая гипотеза верна
- 2) уровень значимости статистики
- 3) вероятность получения наблюдаемого результата, если нулевая гипотеза верна
- 4) разница между наблюдаемыми и теоретическими данными

144. При уровне значимости $\alpha = 0,05$

- 1) мы допускаем 5% вероятность ошибки второго рода
- 2) мы отвергаем нулевую гипотезу, если р-значение меньше 0,05
- 3) мы принимаем нулевую гипотезу, если р-значение меньше 0,05
- 4) мы всегда отвергаем альтернативную гипотезу

145. Ошибка первого рода - это

- 1) ошибка, когда отвергается верная нулевая гипотеза
- 2) ошибка, когда принимается неверная альтернативная гипотеза
- 3) ошибка, когда верная гипотеза принимается за ложную
- 4) ошибка, когда отвергается неверная гипотеза

146. Критерий, который является тестом для сравнения двух выборок называется

- 1) t-критерий для одной выборки
- 2) t-критерий для двух независимых выборок
- 3) критерий хи-квадрат
- 4) z-тест

147. Мощность статистического теста характеризует

- 1) вероятность того, что мы примем нулевую гипотезу
- 2) вероятность того, что мы отвергнем нулевую гипотезу, когда она ложна
- 3) вероятность ошибки первого рода
- 4) вероятность ошибки второго рода

148. Если р-значение больше уровня значимости, то:

- 1) нулевая гипотеза отвергается
- 2) нулевая гипотеза принимается
- 3) невозможно принять решение

- 4) альтернативная гипотеза принимается
149. Альтернативная гипотеза – это
- 1) гипотеза, которую нужно доказать
 - 2) гипотеза, которая утверждает, что данные не равны какой-то постоянной величине
 - 3) гипотеза, противоречащая нулевой гипотезе
 - 4) гипотеза, которая всегда отвергается
150. Критерий хи-квадрат используется
- 1) для проверки значимости разницы между средними
 - 2) для проверки гипотезы о независимости категориальных переменных
 - 3) для проверки нормальности распределения данных
 - 4) для проверки гипотезы о равенстве медиан
151. Для проверки нормальности распределения данных применяется тест
- 1) t-критерий
 - 2) критерий Шапиро-Уилка
 - 3) критерий Фишера
 - 4) критерий Мангена
152. Если ошибка второго рода слишком велика, то
- 1) статистический тест становится менее чувствительным
 - 2) нулевая гипотеза отклоняется
 - 3) получена высокая мощность теста
 - 4) альтернативная гипотеза принимается
153. Процесс выбора гипотезы для тестирования называется
- 1) формулировка гипотез
 - 2) верификация гипотез
 - 3) подтверждение гипотез
 - 4) подбор гипотез
154. Для проверки гипотезы о доли признака применяется критерий
- 1) t-критерий для одной выборки
 - 2) z-тест
 - 3) критерий Фишера
 - 4) ANOVA
155. Метод бутстрапа используется
- 1) для оценки параметров модели
 - 2) для проверки гипотез
 - 3) для оценки точности статистик через повторные выборки
 - 4) для вычисления среднеквадратичного отклонения
156. t-тест для независимых выборок проверяет гипотезу
- 1) нулевая гипотеза: средние значения двух выборок равны
 - 2) нулевая гипотеза: медианы двух выборок равны
 - 3) нулевая гипотеза: дисперсии двух выборок равны
 - 4) нулевая гипотеза: выборки имеют одинаковое распределение
157. Чтобы снизить вероятность ошибки первого рода необходимо
- 1) увеличить уровень значимости
 - 2) уменьшить уровень значимости
 - 3) увеличить размер выборки
 - 4) уменьшить размер выборки
158. Область отклонения для статистики определяет
- 1) стандартная ошибка
 - 2) критическое значение
 - 3) уровень значимости
 - 4) t-распределение

- тест
159. Для проверки различий между средними для нескольких выборок используется
- 1) односторонний t-тест
 - 2) двусторонний t-тест
 - 3) ANOVA
 - 4) тест на равенство дисперсий
160. Примером ошибки второго рода является
- 1) принятие ложной нулевой гипотезы
 - 2) отказ от истинной нулевой гипотезы
 - 3) принятие ложной альтернативной гипотезы
 - 4) принятие истинной нулевой гипотезы
161. F-критерий проверяет
- 1) равенство средних для двух выборок
 - 2) сравнение дисперсий для двух выборок
 - 3) независимость категориальных переменных
 - 4) равенство медиан для двух выборок
- Задания закрытой формы с **множественным выбором**
162. На мощность статистического теста влияют параметры
- 1) размер выборки
 - 2) среднее значение в выборке
 - 3) параметры распределения
 - 4) число степеней свободы
163. Для теста на равенство дисперсий справедливы утверждения
- 1) используется для двух выборок
 - 2) нулевая гипотеза: дисперсии выборок равны
 - 3) для проверки гипотезы используется t-тест
 - 4) применяется только к числовым данным
164. Факторов, влияющие на мощность статистического теста
- 1) размер выборки
 - 2) уровень значимости
 - 3) корреляция
 - 4) аппроксимация
165. Для анализа зависимых выборок используются тесты
- 1) t-тест для парных выборок
 - 2) t-тест для независимых выборок
 - 3) критерий Вилкоксона
 - 4) критерий хи-квадрат
166. Для анализа данных с помощью ANOVA (дисперсионный анализ) верны утверждения
- 1) применяется для сравнения трех и более групп
 - 2) проверяет гипотезу о равенстве средних в группах
 - 3) для использования ANOVA выборки должны быть нормальными и независимыми
 - 4) ожидается, что дисперсии в группах одинаковы
167. Для проверки нормальности данных могут применяться критерии
- 1) критерий Шапиро-Уилка
 - 2) критерий Колмогорова-Смирнова
 - 3) t-тест
 - 4) критерий хи-квадрат
168. Уменьшить вероятность ошибки первого рода помогут следующие действия
- 1) уменьшение уровня значимости
 - 2) увеличение размера выборки

3) применение более мощного теста

4) увеличение уровня значимости

169. Критерии, которые используются для проверки гипотез о пропорциях

1) z-тест для пропорций

2) t-тест для независимых выборок

3) критерий хи-квадрат для независимости

4) критерий Вилкоксона

170. Для оценки согласованности наблюдаемых и ожидаемых данных в категориальных тестах используются следующие характеристики

1) ряды частот

2) критерий хи-квадрат

3) t-распределение

4) кривые распределения

171. Ошибки, которые связаны с гипотезой тестирования

1) ошибка первого рода

2) ошибка второго рода

3) ошибка третьего рода

4) ошибка четвертого рода

Задания на установление **соответствия**

172. Установить соответствие между типами ошибок и их описанием

1) ошибка первого рода

2) ошибка второго рода

А) принятие ложной нулевой гипотезы

В) принятие ложной альтернативной гипотезы

173. Установить соответствие между методами тестирования гипотез с их назначением

1) t-тест

2) критерий хи-квадрат

3) F-тест

А) проверка гипотезы о равенстве дисперсий

В) проверка гипотезы о независимости категориальных переменных

С) проверка гипотезы о равенстве средних

174. Установить соответствие между методами оценки параметров с их назначением

1) Бутстрап

2) метод максимального правдоподобия

3) метод моментов

А) оценка параметров через максимизацию функции правдоподобия

В) оценка параметров через выборочные моменты

С) оценка параметров через повторные выборки

175. Соотнесите тесты с типом данных, для которых они применяются

1) t-тест для одной выборки

2) критерий хи-квадрат

3) t-тест для двух независимых выборок

А) для категориальных данных

В) для сравнения двух средних

С) для анализа одного выборочного среднего

Задания на установление правильной последовательности

176. Установите правильную последовательность этапов проверки гипотезы:

1. формулировка нулевой и альтернативной гипотез

2. выбор статистики теста

3. принятие или отклонение гипотезы
4. сбор и анализ данных
177. Установите правильную последовательность шагов для проведения t-теста для двух независимых выборок:
1. вычисление t-статистики
 2. формулировка гипотез
 3. определение степени свободы
 4. сравнение p-значения с уровнем значимости

Задание открытой формы

178. Гипотеза, утверждающая, что между выборками нет статистически значимых различий, называется _____ гипотезой.
179. Если p-значение меньше уровня значимости, то нулевая гипотеза _____.
180. Когда мощность теста высока, вероятность ошибки второго рода _____.

Тема 3. Проверка статистических гипотез

Задания уровня «знать и уметь»

Задание закрытой формы с единственным выбором

181. Нулевая гипотеза в исследовании о влиянии нового метода обучения на успеваемость студентов формулируется как
- 1) успеваемость студентов не изменилась после применения нового метода обучения
 - 2) успеваемость студентов изменилась после применения нового метода обучения
 - 3) влияет ли новый метод обучения на успеваемость студентов
 - 4) ничего не влияет на уровень образования
182. Компания тестирует новый рекламный слоган и хочет узнать, увеличились ли продажи. Альтернативная гипотеза формулируется как
- 1) продажи не увеличились после использования нового рекламного слогана.
 - 2) качество слогана влияет на продажи
 - 3) качество слогана не влияет на продажи
 - 4) продажи увеличились после использования нового рекламного слогана.
183. Если уровень значимости теста установлен на 0.05, то вероятность ошибки первого рода соответствует
- 1) 0,95.
 - 2) 95%
 - 3) 5%
 - 4) 10%
184. В медицинском исследовании лекарство признано эффективным, хотя оно на самом деле не оказывает влияния. Это говорит о том, что допущена ошибка
- 1) ошибка первого рода (ложноположительный результат).
 - 2) ошибка второго рода (ложноотрицательный результат).
 - 3) ошибка третьего рода (ложноположительный результат).
 - 4) ошибка четвертого рода (ложноотрицательный результат).
185. Компания утверждает, что новый алгоритм обработки данных не лучше старого, хотя на самом деле он эффективнее. Это говорит о том, что допущена ошибка
- 1) ошибка первого рода (ложноположительный результат)..
 - 2) ошибка второго рода (ложноотрицательный результат).
 - 3) ошибка третьего рода (ложноотрицательный результат).
 - 4) ошибка четвертого рода (ложноотрицательный результат).
186. При проверке гипотезы уровень значимости $\alpha = 0.01$, таким образом процент данных попадает в критическую область

- 1) 1%
- 2) 99%
- 3) 0,99
- 4) 0,10

187. Если исследователь хочет проверить, отличается ли средний рост мужчин от 175 см, то альтернативная гипотеза выглядит следующим образом

- 1) $H_1: \mu \neq 175$
- 2) $H_1: \mu = 175$
- 3) $H_1: \mu \neq 0$
- 4) $H_0: \mu \neq 175$

188. Если вероятность ошибки второго рода составляет 0.2, то мощность теста составляет

- 1) 20%
- 2) 80%
- 3) 0,8
- 4) 1,0

189. Для сравнения средних значений двух малых выборок (менее 30 наблюдений), если дисперсии неизвестны, используется

- 1) F-критерий Фишера
- 2) t-критерий Фишера
- 3) критерий Лапласа
- 4) t-критерий Стьюдента

190. Если в результате статистического теста получено $p = 0.03$ при уровне значимости $\alpha = 0.05$, то принимается решение

- 1) отвергнуть нулевую гипотезу
- 2) отвергнуть альтернативную гипотезу
- 3) принять нулевую гипотезу
- 4) пересчитать p

191. Исследователь проверяет, изменилось ли среднее время загрузки веб-сайта (в секундах), которое ранее составляло 3.2 сек. По выборке из 50 измерений было получено среднее значение 3.5 сек и стандартное отклонение 0.8 сек. Если критическое значение $t = 2.01$, то нулевая гипотеза

- 1) отвергается, так как $t = 2.65 > 2.01$, гипотеза
- 2) принимается, так как $t = 2.65 > 2.01$, гипотеза
- 3) отвергается, так как $3.5 > 2.01$, гипотеза
- 4) отвергается, так как $3.2 > 2.01$, гипотеза

192. Фабрика заявляет, что вес упаковки сока в среднем составляет 1 литр. Контрольная группа из 40 упаковок показала средний вес 0.98 л с дисперсией 0.01. Если использовать уровень значимости 0.01, будет ли доказано, что упаковки недоливают?

- 1) да, нет оснований отвергнуть гипотезу
- 2) да, гипотеза отвергается
- 3) нет, нет оснований отвергнуть гипотезу
- 4) нет, гипотеза отвергается

193. Для выборки из 100 наблюдений среднее значение составило 150, а стандартное отклонение 20. Доверительный интервал на 95% построен как [146.08; 153.92]. Можно ли утверждать, что генеральное среднее равно 140?

- 1) да, гипотеза принимается
- 2) нет, гипотеза принимается
- 3) нет, гипотеза отвергается
- 4) да, гипотеза отвергается

194. Если в ходе эксперимента получено значение $p = 0.07$ при уровне значимости 0.05, то следует принять решение, что

- 1) нет оснований отвергнуть гипотезу
- 2) гипотеза отвергается
- 3) изменить эксперимент
- 4) ничего не делать

195. Две группы студентов писали одинаковый тест. В первой группе средний балл 75, во второй – 78. Выборки независимые, объемы 30 и 35, стандартные отклонения 5 и 6 соответственно. При уровне значимости 0.05 можно утверждать, что

- 1) вторая группа написала тест лучше первой
- 2) вторая группа написала тест хуже первой
- 3) группы написали тест одинаково
- 4) первая группа написала тест лучше второй

196. Компания проверяет, связаны ли предпочтения покупателей по брендам с их возрастом. Проверки гипотезы следует использовать

- 1) критерий шапиро-уилка для проверки нормальности распределения
- 2) критерий хи-квадрат для проверки независимости.
- 3) критерий стьюдента
- 4) критерий фишера

197. Опрос показал, что 52% клиентов удовлетворены сервисом компании. При объеме выборки 500 человек требуется проверить, отличается ли доля удовлетворенных клиентов от 50% при уровне значимости 0.05, используя критерий

- 1) Шапиро-Уилка
- 2) хи-квадрат
- 3) Стьюдента
- 4) z-критерий

198. В эксперименте две группы людей тестировали новую программу. Среднее время выполнения задачи в первой группе 12.5 мин, во второй 11.8 мин. Стандартные отклонения 1.5 и 1.2 мин, объемы выборок 30 и 35 соответственно. Можно ли утверждать, что новая программа ускоряет работу с вероятностью 0,95?

- 1) да, с уровнем значимости 0.05 можно утверждать, что программа ускоряет работу
- 2) нет, с уровнем значимости 0.05 можно утверждать, что программа не ускоряет работу
- 3) да, на всех допустимых уровнях значимости можно утверждать, что программа ускоряет работу
- 4) нет, на всех допустимых уровнях значимости можно утверждать, что программа не ускоряет работу

199. Если исследователь желает минимизировать вероятность ошибки первого рода, то ему необходимо выбрать значение уровня значимости

- 1) 0,95 или ниже
- 2) 0,99 или ниже
- 3) 0,01 или ниже
- 4) 0,05 или выше

200. В исследовании сравнивается эффективность двух методов обучения. Первая группа из 25 студентов училась по традиционной методике, а вторая группа из 30 студентов – по новой интерактивной программе. После тестирования оказалось, что средний балл в первой группе составил 68 со стандартным отклонением 7, а во второй группе – 72 со стандартным отклонением 6. Можно ли утверждать, что новая интерактивная программа дает статистически значимое улучшение результатов при уровне значимости 0.05?

- 1) нет, новая программа не дает значимое улучшение ($p < 0.05$)
- 2) нет, новая программа не дает значимое улучшение ($p < 0.50$)
- 3) да, новая программа дает значимое улучшение ($p < 0.05$)
- 4) да, новая программа дает значимое улучшение ($p > 0.50$)

Задания закрытой формы с множественным выбором

201. В эксперименте сравнивается эффективность двух препаратов. В первой группе ($n=30$) среднее время выздоровления 10 дней ($s=2$), во второй ($n=35$) – 9 дней ($s=1.5$). При уровне значимости $\alpha=0,05$, $p\text{-value} = 0,012$ верны следующие утверждения

- 1) различие статистически значимо ($p < 0.05$)
- 2) для проверки подходит критерий Стьюдента
- 3) нулевая гипотеза принимается
- 4) достаточно критерия хи-квадрат

202. При проверке гипотезы о равенстве средних ($\mu_1=\mu_2$) получено $p=0.03$ верны следующие утверждения

- 1) при $\alpha=0.05$ нулевая гипотеза принимается
- 2) при $\alpha=0.01$ нулевая гипотеза отвергается
- 3) p -значение всегда должно быть больше α
- 4) ошибка первого рода невозможна

203. Фабрика производит упаковки крупы с заявленным средним весом 500 г. Контрольная выборка из 25 упаковок показала средний вес 495 г со стандартным отклонением 10 г. При уровне значимости $\alpha = 0.05$ выбрать верные утверждения

- 1) так как $|t|=2.5 > t_{\text{крит}}=2.064$, нулевая гипотеза принимается
- 2) так как $|t|=2.5 > t_{\text{крит}}=2.064$, нулевая гипотеза отвергается
- 3) средний вес упаковок статистически значимо отличается от заявленного
- 4) средний вес упаковок не отличается от заявленного

204. В выборке из 40 наблюдений среднее значение равно 25, а стандартное отклонение 5. При проверке гипотезы о том, что генеральное среднее равно 27 на уровне значимости 0.05, значение t -статистики и вывод о нулевой гипотезе будут следующими

- 1) $t = -3,16$
- 2) $t = 3,16$
- 3) H_0 отвергается
- 4) Нет оснований отвергнуть гипотезу H_0

205. В магазине 52% покупателей совершают повторную покупку. Маркетологи утверждают, что доля постоянных клиентов составляет не менее 55%. Таким образом значение критерия и вывод по гипотезе на уровне значимости 0.05 будут следующими

- 1) $Z = 1,5$
- 2) $Z = -1,5$
- 3) H_0 отвергается
- 4) нет оснований отвергнуть гипотезу H_0

206. Если выборочное среднее составляет 120, стандартное отклонение 15, объем выборки 100. Определить 95%-ный доверительный интервал и проверить гипотезу о том, входит ли в него значение 125.

- 1) [117; 123]
- 2) [135; 163]
- 3) H_0 отвергается
- 4) нет оснований отвергнуть гипотезу H_0

207. Компания утверждает, что дисперсия производственного брака не превышает 4. В выборке из 50 единиц продукции дисперсия составила 5.1. Рассчитать критерий и проверить гипотезу при $\alpha=0.05$.

- 1) $\chi^2=62.88$
- 2) $\chi^2=162.88$
- 3) H_0 отвергается
- 4) нет оснований отвергнуть гипотезу H_0

208. Завод утверждает, что средний вес детали составляет не менее 200 г. В выборке из 36 деталей средний вес составил 198 г, стандартное отклонение 5 г. Рассчитать значение t-критерия и сделать вывод относительно гипотезы при $\alpha=0.05$.

- 1) $t = 2.16$
- 2) $t = -2.40$
- 3) H_0 отвергается
- 4) нет оснований отвергнуть гипотезу H_0

209. Средний балл студентов на экзамене, согласно утверждению преподавателя, составляет 75. В выборке из 50 студентов средний балл оказался 72, стандартное отклонение 10. Требуется рассчитать критерий и проверить гипотезу при $\alpha=0.05$.

- 1) $t = -2.12$
- 2) $t = 3.15$
- 3) H_0 отвергается
- 4) нет оснований отвергнуть гипотезу H_0

210. В городе А 60% жителей пользуются общественным транспортом, а в городе В – 55%. Объем выборки в каждом городе составил 400 человек. Требуется рассчитать критерий и проверить гипотезу при $\alpha=0.05$.

- 1) $Z = 1.82$
- 2) $Z = 4.12$
- 3) H_0 отвергается
- 4) нет оснований отвергнуть гипотезу H_0

Задания на установление **соответствия**

211. Установите соответствие между статистическими критериями и их назначением:

- 1) проверка гипотезы о равенстве дисперсий
- 2) проверка зависимости категориальных переменных
- 3) проверка гипотезы о равенстве средних значений
- 4) проверка гипотезы о пропорциях
- A) t-критерий
- B) Хи-квадрат тест
- C) F-критерий
- D) Z-критерий

212. Установите соответствие между гипотезами и возможными выводами:

- 1) доказано значимое различие
- 2) недостаточно оснований для отклонения
- 3) альтернативная гипотеза справедлива
- 4) альтернативная гипотеза не подтверждена
- A) H_0 отвергается
- B) H_0 принимается
- C) H_1 подтверждается
- D) H_1 отвергается

213. Установите соответствие между видами ошибок и их последствиями:

- 1) ошибка первого рода
- 2) ошибка второго рода
- 3) повышение уровня значимости (α)
- 4) уменьшение уровня значимости (α)
- A) увеличение вероятности ошибки второго рода
- B) принята ложная нулевая гипотеза
- C) увеличение вероятности ошибки первого рода
- D) отклонена истинная нулевая гипотеза

214. Установите соответствие между методами проверки гипотез и их характеристиками:

- 1) параметрические тесты
 - 2) непараметрические тесты
 - 3) одновыборочные тесты
 - 4) двухвыборочные тесты
 - A) требуют нормального распределения данных
 - B) проверяют гипотезу относительно одной выборки
 - C) используются для порядковых и категориальных данных
 - D) сравнивают две независимые или зависимые выборки
- Задания на **установление правильной последовательности**

215. Расположите этапы проверки статистической гипотезы в правильном порядке:

1. выбор уровня значимости (α)
2. формулировка нулевой (H_0) и альтернативной (H_1) гипотез
3. расчет статистического критерия
4. принятие решения о гипотезе

216. Расположите шаги расчета доверительного интервала в правильном порядке:

1. вычисление стандартной ошибки среднего
2. определение уровня значимости (α)
3. определение критического значения (z или t)
4. вычисление границ доверительного интервала

217. Расположите шаги применения критерия Стьюдента для двух независимых выборок в правильном порядке:

1. проверка равенства дисперсий выборок (тест Фишера)
2. вычисление t -статистики
3. проверка нормальности данных (например, тестом Шапиро–Уилка)
4. сравнение с критическим значением и интерпретация результата

Задание **открытой формы**

218. В выборке из 10 значений среднее арифметическое равно 15. Если одно из значений выборки увеличить на 5, то среднее значение изменится на _____.

219. В ходе тестирования гипотезы уровень значимости установлен на 5% ($\alpha=0.05$). Это означает, что вероятность ошибочно отвергнуть нулевую гипотезу составляет _____ (ответ в процентах).

220. В выборке из 100 наблюдений 60 человек выбрали вариант А, а 40 человек — вариант В. Проверяется гипотеза о том, что вероятность выбора варианта А составляет 50%. Таким образом эмпирическое значение доли выборки составляет _____.

Тема 4. Сравнение и создание групп

Задания уровня «знать»

Задание закрытой формы с **единственным выбором**

221. Для сравнения средних значений двух независимых групп используется критерий

- | | | |
|-------------|------------|--------------|
| 1) | | Манна-Уитни |
| 2) | χ^2 | (хи-квадрат) |
| 3) | t-критерий | Стьюдента |
| 4) Фридмана | | |

222. Анализ ANOVA

- 1) используется для сравнения двух групп
- 2) применяется только для несвязанных выборок
- 3) позволяет сравнивать более двух групп
- 4) не требует нормального распределения

223. Для сравнения распределений двух независимых выборок, если данные не соответствуют нормальному распределению применяется тест

- 1) t-критерий Стьюдента
- 2) Манна–Уитни
- 3) дисперсионный анализ (ANOVA)
- 4) критерий Левене

224. Критерий Краскала-Уоллиса является аналогом параметрического теста

- 1) ANOVA
- 2) t-критерия для независимых выборок
- 3) критерия Фишера
- 4) критерия Левене

225. Критерий, который используется для проверки однородности дисперсий

- 1) Левене
- 2) Краскала-Уоллиса
- 3) Манна-Уитни
- 4) t-критерий Стьюдента

226. Статистический метод, который используется для случайного разделения выборки на группы

- 1) кластерный анализ
- 2) дисперсионный анализ
- 3) метод случайной выборки
- 4) регрессионный анализ

227. Критерий, который используется для сравнения пропорций между двумя группами

- 1) χ^2 -критерий
- 2) t-критерий Стьюдента
- 3) дисперсионный анализ
- 4) Манна-Уитни

228. Основным показателем различий между группами в дисперсионном анализе является

- 1) среднее значение
- 2) F-статистика
- 3) медиана
- 4) стандартное отклонение

229. Критерий, который можно использовать для проверки нормальности распределения данных

- 1) Шапиро-Уилка
- 2) Левене
- 3) χ^2 -критерий
- 4) Манна-Уитни

Задания закрытой формы с **множественным выбором**

230. Предположения, которые необходимо проверить перед применением t-критерия Стьюдента

- 1) нормальное распределение данных
- 2) равенство дисперсий

- 3) линейная зависимость между переменными
- 4) независимость наблюдений
231. Методы, которые можно использовать для проверки нормальности распределения
- 1) Шапиро-Уилка
- 2) Колмогорова-Смирнова
- 3) Левене
- 4) Фишера
232. Условия, которые должны быть выполнены для использования дисперсионного анализа (ANOVA)
- 1) группы должны быть независимыми
- 2) данные должны иметь нормальное распределение
- 3) дисперсии в группах должны быть равными
- 4) выборки должны быть одинакового размера
233. К параметрическим критериям относятся
- 1) t-критерий Стьюдента
- 2) критерий Манна-Уитни
- 3) дисперсионный анализ (ANOVA)
- 4) критерий Краскела-Уоллиса
234. Методы, которые применяются для многократного сравнения групп после ANOVA
- 1) тест Тьюки
- 2) тест Бонферрони
- 3) тест Манна-Уитни
- 4) t-критерий для независимых выборок
235. Параметры, которые используются для описания различий между группами
- 1) среднее значение
- 2) медиана
- 3) квартили
- 4) коэффициент корреляции
236. Методы, которые можно использовать для уменьшения влияния выбросов в данных
- 1) использование медианы вместо среднего
- 2) логарифмирование данных
- 3) удаление выбросов
- 4) применение t-критерия Стьюдента
237. Методы для сравнения нескольких групп в непараметрическом анализе
- 1) критерий Краскела-Уоллиса
- 2) ANOVA
- 3) критерий Фридмана
- 4) критерий Манна-Уитни
238. Ошибки, которые могут возникать при проведении множественных сравнений
- 1) ошибка первого рода (ложноположительный результат)
- 2) ошибка второго рода (ложноотрицательный результат)
- 3) ошибка третьего рода (неправильная интерпретация)
- 4) ошибка Левене
239. Непараметрические критерии
- 1) Манна-Уитни
- 2) Краскела-Уоллиса
- 3) Дисперсионный анализ (ANOVA)
- 4) Фридмана
240. Методы для группировки данных

- | | | |
|----|---------------------------|--------|
| 1) | кластерный | анализ |
| 2) | факторный | анализ |
| 3) | регрессионный | анализ |
| 4) | иерархическая агломерация | |
241. Утверждения верные относительно парного t-критерия
- | | | | | | |
|----|-----------------------------------|-------------|---------------|-----------|-----------|
| 1) | используется | для | сравнения | зависимых | выборок |
| 2) | требуется | нормального | распределения | | разностей |
| 3) | используется | для | независимых | | выборок |
| 4) | является непараметрическим тестом | | | | |

Задания на установление **соответствия**

242. Установите соответствие между статистическими критериями и их применением:

- 1) t-критерий Стьюдента
 - 2) Критерий Манна-Уитни
 - 3) ANOVA
 - 4) χ^2 -критерий
- A) сравнение средних двух групп с нормальным распределением
 - B) анализ частотных данных
 - C) сравнение средних более чем двух групп
 - D) сравнение медианных значений двух независимых групп

243. Установите соответствие между тестами и типами данных, к которым они применяются:

- 1) критерий Шапиро-Уилка
 - 2) критерий Левене
 - 3) критерий Краскела-Уоллиса
 - 4) парный t-критерий
- A) сравнение двух зависимых выборок
 - B) проверка равенства дисперсий
 - C) непараметрический аналог ANOVA
 - D) проверка нормальности данных
- Задания на установление **правильной последовательности**

244. Расположите этапы дисперсионного анализа в правильном порядке:

1. формулирование гипотезы
2. интерпретация результатов
3. проверка условий применения метода
4. расчет F-статистики

245. Упорядочите шаги выполнения критерия Манна-Уитни:

1. вычисление суммы рангов для каждой группы
2. объединение выборок и их ранжирование
3. расчет U-статистики
4. сравнение U-статистики с критическим значением

Задание **открытой формы**

246. Если две выборки зависят друг от друга, для их сравнения применяется парный ____.

247. Основная статистика (критерий), рассчитываемая в дисперсионном анализе, – это ____.

248. В статистике гипотеза, которая предполагает отсутствие различий между группами, называется ____.

Тема 4. Сравнение и создание групп

Задания уровня «знать и уметь»

Задание закрытой формы с единственным выбором

249. Две группы студентов пишут экзамен. В первой группе 20 человек, во второй – 30. В этой группе средний балл будет более устойчивым к выбросам

- 1) во второй группе (большой объем выборки снижает влияние выбросов)
- 2) во первой группе (меньший объем выборки снижает влияние выбросов)
- 3) в обоих
- 4) ни в одной

250. В эксперименте две группы испытуемых получили разные методики обучения. В первой группе средний балл – 75, во второй – 80. Утверждается, что вторая методика эффективнее

- 1) нет, не указана ошибка выборки
- 2) да, исключена возможность изменить выборку, испытуемые всегда постоянны
- 3) нет, нужно учитывать дисперсию, размер выборок и проверять статистическую значимость.
- 4) да, во второй группе средний балл больше

251. В исследовании сравниваются две группы: мужчины и женщины. Для проверки различий в среднем росте необходимо использовать критерий

- 1) Пирсона
- 2) Шапиро-Уилка
- 3) t-критерий Стьюдента для независимых выборок.
- 4) Фишера

252. В компании тестируют два вида рекламы и хотят определить, какая из них эффективнее. Для этого необходимо использовать метод

- 1) дисперсионный анализ
- 2) регрессионный анализ
- 3) корреляционный анализ.
- 4) A/B-тестирование

253. В выборке студентов сравнивают уровень стресса в начале и конце семестра у одних и тех же людей. Для проверки гипотезы необходимо использовать статистический тест

- 1) Фишера
- 2) Тьюки
- 3) парный t-критерий Стьюдента.
- 4) критерий Пирсона

254. В эксперименте принимают участие три группы испытуемых. При проверке гипотез для сравнения их средних значений необходимо использовать критерий

- 1) ANOVA (дисперсионный анализ)
- 2) Фишера
- 3) Стьюдента.
- 4) Пирсона

255. В магазине проверяют, отличается ли средний чек покупателей в будни и выходные. Для проверки гипотезы используется критерий

- 1) t-критерий для независимых выборок
- 2) F-критерий для независимых выборок

- 3) Р-критерий для независимых выборок
 4) W-критерий для независимых выборок

256. Метод группировки данных, который чаще всего используют для выявления кластеров в больших данных

- 1) метод сводки
 2) метод k-средних (k-means clustering)
 3) метод Manhattan city block
 4) метод валидации данных

257. Компания тестирует три варианта оформления интернет-магазина. Метод, который поможет определить, есть ли значимые различия в конверсии между вариантами

- 1) метод регрессионного анализа
 2) метод k-средних (k-means clustering)
 3) дисперсионный анализ (ANOVA)
 4) метод валидации данных

258. Исследователь хочет разбить клиентов банка на группы по их поведению при использовании кредитных карт. Метод, который можно использовать для группировки

- 1) k-средних (k-means clustering)
 2) дисперсионного анализа
 3) Manhattan city block
 4) валидации данных

Задания закрытой формы с **множественным выбором**

259. Две группы сотрудников прошли обучение. В первой группе ($n=12$) средняя оценка за тест – 75, стандартное отклонение – 10. Во второй группе ($n=15$) средняя оценка – 80, стандартное отклонение – 12. Используя t-критерий Стьюдента для независимых выборок, проверьте значимость различий на уровне $\alpha=0.05$

- 1) расчетное значение $t=-1.17$, критическое значение $t_{кр}=2.00$
 2) расчетное значение $t=1.98$, критическое значение $t_{кр}=2.00$
 3) гипотеза о равенстве средних не отвергается.
 4) гипотеза о равенстве средних отвергается.

260. В исследовании анализируются две группы студентов. В первой группе ($n=20$) средний рост – 172 см, стандартное отклонение – 8 см. Во второй группе ($n=25$) средний рост – 178 см, стандартное отклонение – 10 см. Проверьте, различаются ли средние показатели на уровне значимости $\alpha=0.01$, используя t-критерий.

- 1) расчетное значение $t=-1.17$, критическое значение $t_{кр}=2.704$
 2) расчетное значение $t=-2.32$, критическое значение $t_{кр}=2.704$
 3) гипотеза о равенстве средних отвергается.
 4) гипотеза о равенстве средних не отвергается.

261. Два рекламных баннера тестировались в разных регионах. В первом регионе ($n=100$) конверсия составила 12%, во втором ($n=120$) – 15%. Проверьте значимость различий, используя критерий разности долей на уровне $\alpha=0.05$.

- 1) расчетное значение $Z=-1.02$, критическое значение $Z_{кр}=1.96$
 2) расчетное значение $Z=-2.32$, критическое значение $Z_{кр}=1.96$
 3) различия значимы.
 4) различия незначимы.

262. Две команды продавцов используют разные стратегии. Средний объем продаж первой команды ($n=30$) – 120 тыс. руб. со стандартным отклонением 15 тыс. руб. Во второй ($n=35$) – 135 тыс. руб. со стандартным отклонением 20 тыс. руб. Проверьте значимость различий t-критерием при $\alpha=0.05$.

- 1) расчетное значение $t=-1.17$, критическое значение $t_{кр}=2.00$
 2) расчетное значение $t=-3.02$, критическое значение $t_{кр}=2.00$

- 3) различие значимо.
- 4) различие незначимо.
263. Для оценки качества обучения сравниваются две группы студентов. В первой группе ($n=40$) 30 человек сдали экзамен, во второй ($n=50$) – 45 человек. Проверьте значимость различий в доле сдавших с помощью критерия разности долей при $\alpha=0.05$.
- 1) расчетное значение $Z=-1.02$, критическое значение $Z_{кр}=1.96$
- 2) расчетное значение $Z=-1.49$, критическое значение $Z_{кр}=1.96$
- 3) различия незначимы.
- 4) различия значимы.

Задания на установление **соответствия**

264. Соотнесите статистические критерии с их назначением:

- 1) t-критерий Стьюдента
- 2) U-критерий Манна–Уитни
- 3) критерий χ^2 (хи-квадрат)
- 4) ANOVA (дисперсионный анализ)
- A) сравнение средних значений двух независимых групп
- B) сравнение двух независимых выборок по ранговым данным
- C) проверка независимости категориальных переменных
- D) сравнение средних значений более чем двух групп

265. Соотнесите описания экспериментов с подходящими статистическими методами:

- 1) исследование эффективности двух методов обучения у студентов
- 2) анализ различий в частоте заболеваний между группами людей с разными привычками
- 3) сравнение трех рекламных стратегий по среднему уровню продаж
- 4) оценка различий в медианном доходе двух профессиональных групп
- A) U-критерий Манна–Уитни
- B) ANOVA (дисперсионный анализ)
- C) критерий χ^2 (хи-квадрат)
- D) t-критерий Стьюдента

266. Соотнесите статистические параметры с их интерпретацией:

- 1) $p\text{-value} < 0.05$
- 2) стандартное отклонение
- 3) доверительный интервал 95%
- 4) размер выборки (n)
- A) мера разброса значений выборки
- B) гипотеза H_0 отвергается, различия значимы
- C) количество наблюдений в исследовании
- D) диапазон, в котором с высокой вероятностью лежит истинное среднее значение

Задания на установление **правильной последовательности**

267. Определите правильную последовательность этапов проверки статистической гипотезы:

1. вычисление статистического критерия
2. определение уровня значимости α
3. формулирование нулевой и альтернативной гипотез
4. сравнение полученного значения критерия с критическим значением

5. принятие решения о гипотезе

268. Расположите этапы проведения дисперсионного анализа (ANOVA) в правильном порядке:

1. вычисление внутригрупповой и межгрупповой дисперсии
2. формулирование гипотез о равенстве средних значений
3. определение критического значения F -критерия
4. вычисление фактического значения F -критерия
5. принятие решения о значимости различий

269. Определите последовательность действий при сравнении двух групп с использованием t -критерия Стьюдента:

1. определение среднего и стандартного отклонения в выборках
2. формулирование гипотез о равенстве средних значений
3. вычисление значения t -критерия
4. сравнение t -критерия с критическим значением
5. принятие решения о гипотезе

Задание открытой формы

270. При проведении t -теста для двух независимых выборок проверяется гипотеза о равенстве _____ в двух группах.

271. Критерий Манна-Уитни используется для сравнения двух _____ выборок, когда данные не соответствуют нормальному распределению.

272. Если p -value меньше уровня значимости α , то нулевая гипотеза _____.

Тема 5. Исследование статистических взаимосвязей

Задания уровня «знать»

Задание закрытой формы с единственным выбором

273. Коэффициент корреляции, который используется для измерения линейной связи между двумя количественными переменными

- | | | | |
|----|--------------------------|------------|----------|
| 1) | коэффициент | корреляции | Спирмена |
| 2) | коэффициент | корреляции | Кендалла |
| 3) | коэффициент | корреляции | Пирсона |
| 4) | коэффициент эластичности | | |

274. Значение коэффициента корреляции Пирсона указывает на отсутствие связи между переменными

- | | |
|--------|----|
| 1) | -1 |
| 2) | 0 |
| 3) | 1 |
| 4) 0.5 | |

275. Для выявления линейных взаимосвязей между переменными используется метод

- 1) регрессионного анализа
- 2) корреляционного анализа

- 3) сплайн-интерполяции
- 4) дисперсионного анализа

276. Коэффициент, который используется для измерения связи между порядковыми переменными

- 1) коэффициент корреляции Пирсона
- 2) коэффициент корреляции Спирмена
- 3) дисперсионный анализ
- 4) регрессионный анализ

277. Коэффициент детерминации (R^2) в регрессионном анализе показывает

- 1) степень зависимости одной переменной от другой
- 2) доля объясненной изменчивости зависимой переменной
- 3) коэффициент наклона регрессионной линии
- 4) стандартное отклонение ошибки

278. Метод, который используется для проверки значимости коэффициента корреляции Пирсона

- 1) t-тест
- 2) χ^2 -критерий
- 3) Z-тест
- 4) критерий Манна-Уитни

279. Коэффициент детерминации $R^2 = 0,85$ в линейной регрессии означает, что

- 1) 85% вариации зависимой переменной объясняется независимой переменной
- 2) 85% наблюдений правильно классифицировано
- 3) 85% коэффициентов регрессии значимы
- 4) 85% вероятности ошибки первого рода

280. Метод, который используется для изучения причинно-следственных связей между переменными

- 1) корреляционный анализ
- 2) регрессионный анализ
- 3) кластерный анализ
- 4) дисперсионный анализ

281. Тип регрессии, который используется, если зависимая переменная является бинарной

- 1) линейная регрессия
- 2) логистическая регрессия
- 3) полиномиальная регрессия
- 4) дисперсионный анализ

282. Статистический метод, который используется для выявления скрытых факторов в данных

- 1) кластерный анализ
- 2) факторный анализ
- 3) регрессионный анализ
- 4) корреляционный анализ

283. Если коэффициент корреляции Пирсона составляет 0,99, то это означает, что между переменными существует

- 1) слабая положительная связь
- 2) сильная положительная связь
- 3) слабая отрицательная связь
- 4) полное отсутствие связи

284. Метод, который используется для выявления связи между двумя категориальными переменными

- 1) коэффициент корреляции Пирсона
- 2) χ^2 -критерий

- 3) линейная регрессия
- 4) коэффициент детерминации
285. Показатель характеризует силу и направление связи между двумя количественными переменными
- 1) коэффициент корреляции
- 2) медиана
- 3) дисперсия
- 4) стандартное отклонение
286. Если две переменные связаны нелинейной зависимостью, то используется коэффициент корреляции
- 1) коэффициент корреляции пирсона
- 2) коэффициент корреляции спирмена
- 3) коэффициент детерминации
- 4) χ^2 -критерий
287. Метод для оценки влияния нескольких независимых переменных на зависимую
- 1) линейная регрессия
- 2) дисперсионный анализ
- 3) множественная регрессия
- 4) корреляционный анализ
- Задания закрытой формы с **множественным выбором**
288. Коэффициенты, относящиеся к корреляционному анализу
- 1) коэффициент корреляции Спирмена
- 2) коэффициент корреляции Пирсона
- 3) коэффициент детерминации
- 4) коэффициент Кендалла
289. Ошибки, которые могут возникнуть при анализе взаимосвязей между переменными
- 1) ошибка первого рода
- 2) ошибка второго рода
- 3) ошибка третьего рода
- 4) ошибка мультиколлинеарности
290. Факторы, которые могут повлиять на значение коэффициента корреляции
- 1) выбросы в данных
- 2) размер выборки
- 3) линейность связи между переменными
- 4) порядок наблюдений в данных
291. Верные утверждения для коэффициента корреляции Спирмена
- 1) коэффициент оценивает силу монотонной зависимости
- 2) используется для категориальных данных
- 3) является параметрическим тестом
- 4) может применяться к ранговым данным
292. Условия, которые должны выполняться для использования коэффициента корреляции Пирсона
- 1) линейная зависимость между переменными
- 2) гомоскедастичность
- 3) переменные должны быть категориальными
- 4) переменные должны быть количественными
293. Для визуализации статистических взаимосвязей чаще используют
- 1) диаграмма рассеяния
- 2) ящик с усами (boxplot)

- 3) тепловая карта корреляций
- 4) столбчатая диаграмма
- 294.Предпосылки, необходимые для построения линейной регрессии
- 1) независимость наблюдений
- 2) гомоскедастичность
- 3) категориальный характер зависимой переменной
- 4) отсутствие мультиколлинеарности
- 295.Методы, используемые для выявления мультиколлинеарности
- 1) коэффициент детерминации
- 2) VIF (Variance Inflation Factor)
- 3) корреляционная матрица
- 4) хи-квадрат тест
- 296.Коэффициенты корреляции, которые используются для порядковых переменных
- 1) коэффициент Пирсона
- 2) коэффициент Спирмена
- 3) коэффициент Кендалла
- 4) коэффициент Детерминации
- 297.Методы, позволяющие проверить статистическую значимость коэффициента корреляции
- 1) t-критерий Стьюдента
- 2) χ^2 -критерий
- 3) бутстрап-методы
- 4) F-критерий
- 298.Проблемы, которые могут возникнуть при использовании линейной регрессии
- 1) мультиколлинеарность
- 2) гетероскедастичность
- 3) нелинейность данных
- 4) высокая размерность данных
- 299.Методы, которые используются для прогнозирования значений на основе имеющихся данных
- 1) линейная регрессия
- 2) логистическая регрессия
- 3) факторный анализ
- 4) деревья решений
- 300.Тесты, которые используются для проверки гипотезы о независимости двух категориальных переменных
- 1) χ^2 -критерий
- 2) F тест
- 3) коэффициент корреляции Спирмена
- 4) точный критерий Фишера
- 301.Методы, которые используются для выявления скрытых факторов, влияющих на данные
- 1) факторный анализ
- 2) кластерный анализ
- 3) логистическая регрессия
- 4) главные компоненты анализа (РС)
- 302.Методы, которые используются для моделирования нелинейных зависимостей
- 1) полиномиальная регрессия
- 2) логистическая регрессия

- 3) линейная регрессия
4) нейронные сети

Задания на установление **соответствия**

303. Установите соответствие между коэффициентами корреляции и их применением:

- 1) коэффициент корреляции Пирсона
- 2) коэффициент корреляции Спирмена
- 3) коэффициент Кендалла
- 4) коэффициент детерминации (r^2)
- A) доля объясненной вариативности в регрессии
- B) связь между порядковыми переменными
- C) альтернативный метод для ранговых данных
- D) линейная зависимость между количественными переменными

304. Установите соответствие между статистическими методами и их основными задачами:

- 1) корреляционный анализ
- 2) регрессионный анализ
- 3) дисперсионный анализ (ANOVA)
- 4) факторный анализ
- A) прогнозирование значений зависимой переменной
- B) оценка силы связи между переменными
- C) выявление скрытых факторов в данных
- D) сравнение средних значений в группах

305. Установите соответствие между коэффициентами корреляции и их интерпретацией:

- 1) 0,0
- 2) 0,8
- 3) -0,5
- 4) -1,0
- A) полное отсутствие связи
- B) умеренная отрицательная связь
- C) сильная положительная связь
- D) абсолютная обратная связь

306. Соотнесите тип коэффициента корреляции с его областью применения:

- 1) коэффициент Пирсона
- 2) коэффициент Спирмена
- 3) коэффициент Кендалла
- 4) χ^2 -критерий
- A) анализ связи категориальных переменных
- B) порядковые данные с небольшими выборками
- C) порядковые данные или нелинейные зависимости
- D) количественные переменные с линейной связью

307. Установите соответствие между статистическими тестами и их назначением:

- 1) t-тест
- 2) χ^2 -критерий
- 3) F-критерий
- 4) Z-тест

- A) проверка различий между несколькими группами
- B) проверка независимости категориальных переменных
- C) сравнение средних двух групп
- D) оценка параметров нормального распределения

Задания на установление правильной последовательности

308. Упорядочите этапы выполнения корреляционного анализа:

1. интерпретация результатов
2. выбор метода корреляции
3. вычисление коэффициента корреляции
4. проверка предпосылок

309. Расположите этапы корреляционного анализа в правильном порядке:

1. вычисление коэффициента корреляции
2. интерпретация результатов
3. проверка статистической значимости
4. определение типа данных и выбор метода корреляции

310. Упорядочите шаги построения линейной регрессии:

1. определение зависимой и независимой переменной
2. подбор модели и оценка параметров
3. проверка качества модели (r^2 , остатки, значимость)
4. интерпретация результатов и выводы

311. Упорядочите этапы проведения дисперсионного анализа (ANOVA):

1. Формулирование гипотезы о равенстве средних
2. Интерпретация результатов и принятие решения
3. Разделение общей вариации на объясненную и остаточную
4. Вычисление F-статистики

312. Расположите этапы проверки гипотезы в статистическом анализе:

1. вычисление статистики и p-значения
2. определение нулевой и альтернативной гипотез
3. выбор статистического критерия
4. сравнение p-значения с уровнем значимости и вынесение решения

Задание открытой формы

313. Основная цель корреляционного анализа – выявить _____ между переменными.

314. Если коэффициент корреляции равен -1, это означает _____ линейную зависимость.

315. Если между двумя переменными нет статистически значимой связи, то коэффициент корреляции будет близок к ____.

316. В регрессионной модели положительный коэффициент перед фактором означает ____ связь между результативным показателем и фактором.

317. В корреляционном анализе, если увеличение одной переменной сопровождается уменьшением другой, то такая связь называется ____.

Тема 5. Исследование статистических взаимосвязей

Задания уровня «знать и уметь»

Задание закрытой формы с единственным выбором

318. Корреляция между двумя переменными составила $r=0.850$, это означает

- | | | | |
|----|-----------|----------|-------|
| 1) | сильную | прямую | связь |
| 2) | сильную | обратную | связь |
| 3) | умеренную | прямую | связь |
| 4) | умеренную | обратную | связь |

319. Коэффициент детерминации $R^2=0,64$, таким образом, процент вариации зависимой переменной объясняется независимой на

- | | |
|----|-------|
| 1) | 0.34% |
| 2) | 64% |
| 3) | 34% |
| 4) | 0.64% |

320. Если корреляция между двумя переменными $r = -0,415$, то связь

- | | | | |
|----|-----------|----------|-------|
| 1) | сильная | прямая | связь |
| 2) | сильная | обратная | связь |
| 3) | умеренная | прямая | связь |
| 4) | умеренная | обратная | связь |

321. При вычислении корреляции Спирмена получено значение $r = 0$, то связь между переменными

- | | | |
|----|---------|-------------|
| 1) | сильная | прямая |
| 2) | | отсутствует |
| 3) | сильная | обратная |
| 4) | слабая | прямая |

322. Если p -value в корреляционном анализе равно 0,02 при уровне значимости $\alpha=0,05$, то связь

- | | | |
|----|---------------|-------------------|
| 1) | статистически | значима |
| 2) | статистически | не значима |
| 3) | сильная | между переменными |
| 4) | отсутствует | между переменными |

323. В парной линейной регрессии коэффициент при независимой переменной равен 2, что означает

- | | | | |
|----|---|-------------------|---------|
| 1) | при увеличении независимой переменной на 1 единицу зависимая увеличивается на | 2 | единицы |
| 2) | при увеличении зависимой переменной на 1 единицу независимая увеличивается на | 2 | единицы |
| 3) | связь между факторным признаком и результативным | сильная и прямая | |
| 4) | связь между факторным признаком и результативным | слабая и обратная | |

324. Дана модель линейной регрессии: $Y = 5 + 2X$, где Y – прогнозируемая переменная, X – независимая переменная. Определите значение Y , если $X = 4$.

- 1) 28
 2) 13
 3) -13
 4) 18

325. Дана трехфакторная модель линейной регрессии: $Y = 3 + 2X_1 - 0.5X_2 + 4X_3$, где: Y – зависимая переменная, X_1 , X_2 , X_3 – независимые переменные. Определите значение Y , если: $X_1 = 5$, $X_2 = 2$, $X_3 = 3$.

- 1) 28
 2) 29
 3) 24
 4) 19

Задания закрытой формы с **множественным выбором**

326. В таблице приведены данные о количестве часов подготовки к экзамену (X) и набранных баллах (Y) для 5 студентов:

Студент	Часы подготовки (X)	Баллы на экзамене (Y)
1	2	50
2	4	60
3	6	65
4	8	70
5	10	80

Коэффициент корреляции Пирсона между X и Y и Пего значимость при уровне значимости $\alpha = 0.05$ ($t_{\text{критич}} = 2,776$) для выборки

- 1) $r = 0.976$
 2) $r = -0.976$
 3) коэффициент корреляции значим, $t = 7.45$
 4) коэффициент корреляции не значим, $t = 7.45$

327. По следующим данным постройте линейное уравнение регрессии, вычислите линейный коэффициент корреляции: $= 100$, $= 10$, $= 8$, $= 136$, $= 100$, $= 4,8$

- 1) $r = 0.56$
 2) $r = -0.56$
 3) $y = 4,8 - 0,56x$
 4) $y = 4,8 - 0,56x$

328. Компания занимается грузоперевозками и хочет спрогнозировать затраты на топливо (литры) (Y) в зависимости от расстояния поездки в километрах (X). По результатам наблюдений за 10 рейсами была получена модель регрессии: $Y = 50 + 4,4X$. Дать интерпретацию свободного члена и параметра перед фактором

- 1) средние затрат на топливо без учета влияния факторов составляют 50 литров
 2) при увеличении расстояния поездки на 1 км, затраты на топливо увеличиваются в среднем на 4.4 литров.
 3) средние затрат на топливо без учета влияния факторов составляют 4.4 литра
 4) при увеличении расстояния поездки на 4.4 км, затраты на топливо увеличиваются в среднем на 50 литров

329. Даны суммы: $\sum X = 20$, $\sum Y = 40$, $\sum XY = 200$, $\sum X^2 = 120$, $n = 5$. Найти коэффициенты a и b в уравнении линейной регрессии $Y = a + bX$

- 1) $a = 4$
 2) $b = 1$
 3) $a = 5$
 4) $b = -1$

330. Логистическая компания анализирует влияние двух факторов на затраты на топливо (Y): расстояния поездки (X_1 , км) и массы груза (X_2 , тонны). По результатам построенной регрессионной модели получено: $Y = 10 + 0,12X_1 + 2,5X_2$. Коэффициент

множественной корреляции $R = 0,9$. Рассчитать коэффициент детерминации R^2 и интерпретировать его

- 1) $R^2 = 0,81$
- 2) $R^2 = 0,10$
- 3) 81% изменчивости затрат на топливо объясняется расстоянием поездки и массой груза.
- 4) 10% изменчивости затрат на топливо объясняется расстоянием поездки и массой груза.

331. Исследуется зависимость между количеством часов подготовки (X) и результатом теста (Y). Известно: среднее значение $X = 5$ часов, $Y = 70$ баллов; ковариация $Cov(X, Y) = 30$; дисперсия $X = 4$. Найти 1) коэффициент корреляции Пирсона (r), 2) коэффициент наклона (b) в уравнении линейной регрессии $Y = a + bX$.

- 1) $r = 0.75$
- 2) $r = -0.75$
- 3) $b = 9.4$
- 4) $b = 7.5$

332. Дано уравнение регрессии: $Y = 10 + 2X$. Известно, что при $X = 3$ фактическое значение $Y = 18$. Требуется найти: 1) Прогнозируемое значение Y при $X = 5$, 2) Остаток для точки $X = 3$.

- 1) прогнозируемое значение $Y = 18$
- 2) прогнозируемое значение $Y = 20$
- 3) остаток $= 0$
- 4) остаток $= 2$

333. Исследуется зависимость между возрастом сотрудников (X) и их зарплатой (Y) в тыс. руб. Среднее значение $X = 35$ лет, $Y = 80$ тыс. руб., сумма произведений отклонений: $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 420$, сумма квадратов отклонений для X: $\sum(X_i - \bar{X})^2 = 100$. Найти: 1) коэффициент корреляции Пирсона (r). 2) угловой коэффициент (b) в уравнении регрессии $Y = a + bX$

- 1) $r = 0,600$
- 2) $r = 0,421$
- 3) $b = 4,2$
- 4) $b = 2,4$

334. Уравнение регрессии для зависимости продаж (Y) от затрат на рекламу (X) в млн руб.: $Y = 50 + 10X$. Известно, что при $X = 2$ фактический объем продаж составил $Y = 75$ млн руб. Найти: 1) прогнозируемые продажи при $X = 3$, 2) остаток модели для точки $X = 2$.

- 1) прогнозируемые продажи составили 110 млн. руб.
- 2) прогнозируемые продажи составили 80 млн. руб.
- 3) остаток $= -5$ млн. руб.
- 4) остаток $= 5$ млн. руб.

335. Для студентов записаны часы самостоятельной работы (X) и баллы на экзамене (Y): Ковариация $Cov(X, Y) = 24$, Дисперсия $X = 16$, Дисперсия $Y = 25$. Найти: 1) коэффициент корреляции Пирсона (r), 2) свободный член (a) в уравнении регрессии, если $\bar{X} = 10$, $\bar{Y} = 60$.

- 1) $r = 0.600$.
- 2) $r = 0.700$.
- 3) $a = 45$.
- 4) $a = 41$.

Задания на установление соответствия

336. Для выборки данных из 10 наблюдений получены следующие характеристики: сумма $X = 55$, сумма $Y = 120$, сумма $XY = 700$, сумма $X^2 = 385$, сумма $Y^2 = 1500$. Сопоставьте расчетную величину и ее значение:

- 1) коэффициент корреляции (r)
- 2) параметр уравнения регрессии (b)

- 3) константа (a)
- 4) прогноз $\hat{Y}$ при $X=8$
- A) 1,5
- B) 4,5
- C) 16,5
- D) 0,92

337. Дано уравнение регрессии $\hat{Y} = 3 + 2X$. Известны следующие фактические значения: при $X=2$, $Y=8$; при $X=5$, $Y=12$. Сопоставьте расчетную величину и ее значение:

- 1) Прогноз $\hat{Y}$ при $X=2$
- 2) Остаток (ϵ) для $X=2$
- 3) Прогноз $\hat{Y}$ при $X=5$
- 4) Остаток (ϵ) для $X=5$
- A) 7
- B) 13
- C) -1
- D) 1

338. Для анализа связи между двумя переменными получены: ковариация $\text{Cov}(X,Y) = 18$; дисперсия $\sigma_x^2 = 9$; дисперсия $\sigma_y^2 = 36$; среднее $\bar{X} = 10$; среднее $\bar{Y} = 25$. Сопоставьте расчетную величину и ее значение:

- 1) коэффициент корреляции (r)
- 2) параметр уравнения регрессии (b)
- 3) константа (a)
- 4) прогноз $\hat{Y}$ при $X=12$
- A) 29
- B) 5
- C) 2
- D) 0,5

Задания на установление правильной последовательности

339. Установите правильную последовательность шагов при построении линейной регрессионной модели:

1. сбор данных
2. расчет коэффициента корреляции
3. построение диаграммы рассеяния
4. проверка значимости коэффициентов
5. расчет параметров уравнения регрессии
6. интерпретация результатов

340. Расположите в правильном порядке этапы проверки статистической гипотезы на незначимость параметров регрессионного уравнения:

1. формулировка нулевой и альтернативной гипотез
2. выбор уровня значимости
3. расчет тестовой статистики
4. определение критической области
5. сравнение статистики с критическим значением
6. принятие решения о гипотезе

341. Укажите последовательность проверки допущений линейной регрессии:

1. проверка нормальности распределения остатков
2. анализ гомоскедастичности остатков
3. проверка на автокорреляцию
4. построение графиков остатков
5. проверка мультиколлинеарности

Задание открытой формы

342. В уравнении линейной регрессии $Y = 3 + 0.5X$, при увеличении X на 2 единицы, Y увеличится на ____ единиц.

343. Если p -значение равно для параметра регрессионного уравнения перед фактором равно 0.03, а уровень значимости $\alpha = 0.05$, то нулевую гипотезу о незначимости параметра регрессионного уравнения _____.

344. Если коэффициент корреляции Пирсона равен +0,899, это означает, что связь между переменными _____ и тесная.

345. В уравнении регрессии $\hat{Y} = 15 + 3X$ угловой коэффициент показывает, что при увеличении X на 1 единицу Y увеличивается на ____ единиц.

**Тема 6. Актуальность и роль визуализации данных
в аналитике больших данных**

Задания уровня «знать»

Задание закрытой формы с единственным выбором

346. Основная цель визуализации данных

- 1) запутать пользователей
- 2) сделать графики
- 3) облегчить интерпретацию данных
- 4) уменьшить объем данных

347. Тип диаграммы лучше всего использовать для сравнения частей целого

- 1) линейный график
- 2) круговая диаграмма
- 3) гистограмма
- 4) коробчатая диаграмма

348. Инструмент чаще всего используется для интерактивной визуализации данных из предложенных

- 1) Excel
- 2) Data Lens
- 3) Notepad
- 4) Word

349. График лучше всего подходит для отображения временных рядов

- 1) гистограмма
- 2) круговая диаграмма
- 3) линейный график
- 4) точечная диаграмма

350. Метод визуализации данных, который помогает выявить выбросы в данных

- 1) линейный график
- 2) коробчатая диаграмма (boxplot)

3) круговая диаграмма

4) столбчатая диаграмма

351.Метод визуализации, при котором данные отображаются в виде взаимосвязанных узлов

1) диаграмма рассеяния

2) граф (network graph)

3) круговая диаграмма

4) гистограмма

352.Тип диаграммы, который позволяет показать распределение данных

1) точечная диаграмма

2) гистограмма

3) линейный график

4) круговая диаграмма

353.Термин «тепловая карта» в контексте визуализации данных означает

1) карта температур в городе

2) метод отображения плотности данных с использованием цвета

3) график зависимости температуры от времени

4) вид линейного графика

354.Формат данных, который чаще всего используется для визуализации

1) PDF

2) JSON

3) MP3

4) DOCX

355.Термин, который описывает процесс использования графики для представления числовой информации

1) визуализация данных

2) кодирование

3) фильтрация данных

4) агрегация данных

356.Цвет, который чаще всего используется для обозначения негативных значений в визуализации

1) синий

2) красный

3) зеленый

4) желтый

357.Принцип, который важно учитывать при создании визуализации данных

1) использовать максимальное количество цветов

2) упрощать восприятие информации

3) добавлять анимацию в каждый элемент

4) использовать как можно больше текста

358.Диаграмма рассеяния показывает

1) временной ряд

2) взаимосвязь между двумя переменными

3) распределение частот

4) процентное соотношение частей

359.Метод визуализации, который используется для анализа текстовых данных

1) гистограмма

2) облако слов

3) круговая диаграмма

4) столбчатая диаграмма

Задания закрытой формы с множественным выбором

360. Диаграммы, которые подходят для отображения категориальных данных

- 1) гистограмма
- 2) столбчатая диаграмма
- 3) круговая диаграмма
- 4) линейный график

361. Популярные инструменты, которые используются для визуализации данных

- 1) Power BI
- 2) Tableau
- 3) MS Word
- 4) Data Lens

362. Принципы, которые делают визуализацию эффективной?

- 1) простота
- 2) минимизация текста
- 3) использование множества цветов
- 4) четкая структура

363. Важные параметры при выборе типа визуализации

- 1) количество переменных
- 2) тип данных
- 3) размер файла
- 4) цель анализа

364. Элементы, улучшающие восприятие диаграмм

- 1) легенда
- 2) подписи осей
- 3) случайные цвета
- 4) заголовок

365. Диаграммы, которые лучше всего подходят для отображения трендов во времени

- 1) линейный график
- 2) столбчатая диаграмма
- 3) тепловая карта
- 4) гистограмма

366. Преимущества, которые дает интерактивная визуализация данных

- 1) возможность фильтрации данных
- 2) улучшение понимания информации
- 3) отсутствие необходимости в анализе данных
- 4) возможность детального исследования данных

367. Инструменты, которые позволяют строить интерактивные графики

- 1) Power BI
- 2) Tableau
- 3) Matplotlib
- 4) Data Lens

368. Цвета, которые чаще всего используются для передачи эмоционального воздействия в визуализации данных

- 1) красный
- 2) зеленый
- 3) синий
- 4) серый

369. Типы диаграмм, которые помогают сравнивать несколько категорий данных

- 1) столбчатая диаграмма
- 2) линейный график
- 3) круговая диаграмма
- 4) гистограмма

370. Ошибки, которые могут ухудшить восприятие визуализации данных

- 1) слишком много цветов
- 2) перегруженность графика данными
- 3) использование подписи осей
- 4) отсутствие заголовка

371. Элементы, которые важны для улучшения читаемости диаграмм

- 1) четкие подписи осей
- 2) контрастные цвета
- 3) мелкий шрифт
- 4) простая структура

372. Графики, которые помогают анализировать корреляцию между двумя переменными

- 1) диаграмма рассеяния
- 2) линейный график
- 3) гистограмма
- 4) круговая диаграмма

373. Типы визуализации, которые чаще всего используются для работы с геоданными

- 1) тепловая карта
- 2) картограмма
- 3) линейный график
- 4) диаграмма рассеяния

374. Методы визуализации, которые применяются для анализа текстовых данных

- 1) облако слов
- 2) дендрограмма
- 3) гистограмма
- 4) столбчатая диаграмма

Задания на установление **соответствия**

375. Соотнесите типы диаграмм с их основным назначением:

- 1) линейный график
- 2) круговая диаграмма
- 3) гистограмма
- 4) диаграмма рассеяния
- A) анализ временных рядов
- B) анализ взаимосвязи между переменными
- C) отображение распределения данных
- D) сравнение долей в целом

376. Установите соответствие между инструментами визуализации и их основными особенностями:

- 1) Tableau
- 2) Power BI
- 3) Plotly
- A) создание веб-графиков и интерактивных диаграмм
- B) интеграция с продуктами Microsoft для создания интерактивных дашбордов
- C) интерактивные дашборды

377. Соотнесите элементы визуализации с их функцией:

- 1) легенда
- 2) подписи осей
- 3) заголовок
- 4) цветовая шкала

- A) объяснение содержания диаграммы
- B) отображение градиента значений
- C) объяснение цветов и категорий
- D) указание значений шкалы

378.Соотнесите типы визуализаций с областями их применения:

- 1) тепловая карта
- 2) коробчатая диаграмма
- 3) дендрограмма
- 4) столбчатая диаграмма
- A) анализ выбросов и медианы
- B) географическое распределение данных
- C) сравнение категорий
- D) кластерный анализ

379.Соотнесите термины, связанные с визуализацией, с их определениями:

- 1) дашборд
- 2) интерактивная визуализация
- 3) геовизуализация
- 4) инфографика
- A) графическое представление информации в статичном виде
- B) визуализация данных на картах
- C) возможность изменять и фильтровать данные на графике
- D) совокупность визуализаций для мониторинга данных

Задания на **установление правильной последовательности**

380.Расположите шаги построения эффективной визуализации данных в правильном порядке:

1. определение цели визуализации
2. выбор типа графика
3. подготовка данных
4. создание диаграммы
5. интерпретация и анализ

381.Упорядочите этапы работы с интерактивной визуализацией в Power BI:

1. загрузка и подготовка данных
2. публикация дашборда
3. добавление интерактивных фильтров
4. создание визуализаций
5. настройка дизайна

382.Упорядочите шаги анализа временных рядов с помощью визуализации:

1. сбор данных
2. интерпретация и выводы
3. выявление сезонных колебаний
4. очистка данных
5. построение графика тренда

383.Расположите этапы создания дашборда в Tableau в правильном порядке:

1. загрузка данных
2. выбор необходимых полей
3. добавление фильтров и интерактивных элементов
4. создание диаграмм
5. публикация дашборда

384.Расположите этапы обработки геоданных перед их визуализацией:

1. выбор типа картографической визуализации
2. сбор данных (координаты, географические объекты)
3. нанесение данных на карту
4. очистка и обработка данных
5. анализ пространственных закономерностей

Задание **открытой формы**

385.Гистограмма используется для отображения _____ типа переменной

386.Тепловая карта представляет собой визуализацию данных с использованием _____.

387.Основное преимущество визуализации данных – это быстрое выявление _____.

388.Круговая диаграмма используется для сравнения _____.

389.Коробчатая диаграмма (boxplot) помогает определить _____ в данных.

Тема 6. Актуальность и роль визуализации данных в аналитике больших данных

Задания уровня «знать и уметь»

Задание закрытой формы с **единственным выбором**

390.Для сравнения нескольких категорий чаще всего используется

- 1) тепловая карта
- 2) линейный график
- 3) столбчатая диаграмма
- 4) ящик с усами (boxplot)

391.Тип диаграммы, который лучше всего подходит для отображения соотношения частей к целому

- 1) гистограмма
- 2) линейный график
- 3) круговая диаграмма
- 4) диаграмма рассеяния

392.Метод визуализации, позволяющий анализировать корреляцию между двумя переменными

- 1) Boxplot
- 2) диаграмма рассеяния
- 3) круговая диаграмма
- 4) столбчатая диаграмма

393. Формат, в котором чаще всего хранятся данные для визуализации в BI-системах

- 1) .txt
- 2) .json
- 3) .csv
- 4) .docx

394. Визуализация, представляющая данные в виде взаимосвязанных точек и линий

- 1) гистограмма
- 2) линейный график
- 3) круговая диаграмма
- 4) тепловая карта

395. Тип диаграммы который чаще всего используется для отображения распределения данных с выбросами

- 1) линейный график
- 2) столбчатая диаграмма
- 3) круговая диаграмма
- 4) Boxplot (ящик с усами)

Задания закрытой формы с **множественным выбором**

396. Элементы, которые включает в себя boxplot-график

- 1) медиана
- 2) среднее значение
- 3) «усы» (whiskers)
- 4) дисперсия

397. С помощью boxplot можно определить

- 1) наличие выбросов
- 2) размах данных
- 3) точные значения всех наблюдений
- 4) симметричность распределения

398. Если boxplot имеет длинный верхний «ус», то

- 1) данные распределены симметрично
- 2) в верхней части данных больше вариации
- 3) в выборке есть выбросы
- 4) среднее значение больше медианы

399. Статистические характеристики, которые можно определить по boxplot

- 1) квартильное расстояние
- 2) среднее арифметическое
- 3) медиана
- 4) выбросы

400. Если верхний «ус» в boxplot намного длиннее нижнего, то это означает

- 1) в выборке есть выбросы
- 2) данные имеют положительную асимметрию
- 3) данные распределены равномерно
- 4) размах данных одинаковый в обе стороны

401. Особенности характерны для интерактивных инструментов визуализации данных Tableau и Power BI

- 1) возможность фильтрации данных
- 2) генерация отчетов в реальном времени
- 3) возможность обработки данных без подключения к внешним источникам
- 4) автоматическое выявление аномалий в данных

402. Виды диаграмм, которые лучше всего подходят для отображения трендов во временных рядах

- 1) линейный график
 - 2) boxplot
 - 3) точечная диаграмма
 - 4) диаграмма областей
403. Графики, которые подходят для отображения распределения данных
- 1) boxplot
 - 2) гистограмма
 - 3) линейный график
 - 4) круговая диаграмма
404. Программы, позволяющие создавать интерактивные визуализации данных
- 1) Tableau
 - 2) Excel
 - 3) Power BI
 - 4) Google Data Studio
405. Инструменты, которые являются облачными сервисами для визуализации данных
- 1) Google Data Studio
 - 2) Looker
 - 3) Matplotlib
 - 4) Qlik Sense

Задания на установление **соответствия**

406. Сопоставьте BI-системы с их характеристиками.

- 1) Power BI
 - 2) Tableau
 - 3) Google Data Studio
 - 4) Qlik Sense
- A) облачная BI-система от Google
 - B) гибкая система с мощной визуализацией
 - C) инструмент от Microsoft для анализа данных
 - D) использует ассоциативную модель данных

407. Сопоставьте типы графиков с их основным предназначением.

- 1) линейный график
 - 2) гистограмма
 - 3) круговая диаграмма
 - 4) boxplot
- A) показывает соотношение частей к целому
 - B) используется для отображения трендов во времени
 - C) позволяет анализировать распределение данных
 - D) помогает выявить выбросы и медиану данных

408. Сопоставьте виды таблиц с их функциями.

- 1) сводная таблица
 - 2) факт-таблица
 - 3) дашбордная таблица
 - 4) источниковая таблица
- A) хранит исходные данные в неизменном виде
 - B) отображает агрегированные значения по категориям
 - C) используется для интерактивных отчетов
 - D) содержит ключевые показатели и метрики

Задания на установление **правильной последовательности**

409. Расположите этапы работы с BI-системой в правильном порядке:

- 1) визуализация данных
- 2) подключение и загрузка данных
- 3) анализ данных и создание отчетов
- 4) очистка и подготовка данных

410. Определите правильную последовательность действий для создания графика в BI-системе:

- 1) выбор типа диаграммы
- 2) загрузка и подключение данных
- 3) настройка параметров визуализации (цвета, подписи и т. д.)
- 4) выбор полей для осей графика

411. Расположите действия по созданию сводной таблицы в BI-системе в правильном порядке:

- 1) выбор полей для строк и столбцов
- 2) группировка и агрегирование данных
- 3) загрузка данных в систему
- 4) добавление вычисляемых показателей

Задание открытой формы

412. Визуализация данных помогает выявлять закономерности, тренды и _____ в больших объемах информации.

413. График ____ типа используется для отображения трендов во времени и динамики изменений.

414. В BI-системах, таких как Power BI и Tableau, для объединения данных из нескольких источников применяется процесс, называемый _____.

415. В диаграмме типа Boxplot горизонтальная линия внутри прямоугольника обозначает

Краткие методические указания

Тест содержит задания с выбором одного или нескольких ответов.

Шкала оценки

С помощью тестов эффективно проверяется степень владения обучающимися материалом практически всех дисциплин *на уровне «знать»*. При этом проверяется знание понятийного аппарата дисциплины, фактов, законов, закономерностей, формул, норм, правил и т.п., а к оцениванию заданий подобного типа применяется дихотомическая шкала: верно / неверно.

Вместе с тем, для самого широкого круга дисциплин возможна разработка тестовых заданий *на уровне «знать и уметь»*. Решая задания такого типа, обучающийся выбирает один из возможных изученных способов решения стандартных, типовых задач. При оценивании таких заданий учитываются с определенным баллом не только полностью, но и частично выполненные задания (политомическая шкала: верно / частично верно / неверно).

Банк тестовых заданий, построенный по уровневой модели оценивания результатов обучения, состоит из двух блоков, представленных заданиями двух уровней («знать», «знать и уметь»).

Первый блок – задания уровня «знать», выявляющие в основном «знаниевый» компонент содержания дисциплины. Цель тестирования заданиями данного блока состоит в определении уровня усвоения обучающимися базовых понятий, закономерностей, фактов и т.п., относящихся к дисциплине.

Второй блок – задания уровня «знать» и «уметь», позволяющие оценить не только знания по дисциплине, но и умения пользоваться ими при решении стандартных задач.

Цель тестирования заданиями данного блока состоит в проверке возможностей обучающихся использовать полученные знания и умения для выполнения типовых заданий. Во втором блоке задания направлены не столько на оценку определенных знаний, сколько на диагностику более широкого набора умений в рамках целого раздела дисциплины, выходящего за рамки одной темы.

1. 1. Шкала для оценивания

№	Уровень обученности	Показатели оценки результатов обучения
1.	Начальный	Менее 70% баллов за задания каждого из блоков Б.1 и Б.2
2.	Базовый	Не менее 70% баллов за задания блока Б.1 и менее 70% баллов за задания каждого из блока Б.2 или Не менее 70% баллов за задания блока Б.2 и менее 70% баллов за задания каждого из блока Б.1
3.	Повышенный	Не менее 70% баллов за задания каждого из блоков Б.1 и Б.2