

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

КАФЕДРА МАТЕМАТИКИ И МОДЕЛИРОВАНИЯ

Рабочая программа дисциплины (модуля)  
**СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ**

Направление и направленность (профиль)  
41.03.01 Зарубежное регионоведение. Политическая, правовая и деловая культура стран  
Азии

Год набора на ОПОП  
2024

Форма обучения  
очная

Владивосток 2024

Рабочая программа дисциплины (модуля) «Статистический анализ данных» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 41.03.01 Зарубежное регионоведение (утв. приказом Минобрнауки России от 15.06.2017г. №553) 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., доцент, Кафедра математики и моделирования, Kocheva.EV@vvsu.ru*

*Шумик Е.Г., кандидат экономических наук, доцент, Кафедра математики и  
моделирования, Ekaterina.Shumik1@vvsu.ru*

Утверждена на заседании кафедры математики и моделирования от 23.05.2024 ,  
протокол № 9

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Галимзянова К.Н.

<b>ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ</b>	
Сертификат	1599657997
Номер транзакции	000000000D42A41
Владелец	Галимзянова К.Н.

## 1 Цель, планируемые результаты обучения по дисциплине (модулю)

Целью преподавания дисциплины «Прикладной статистический анализ данных» является углубление знаний прикладной математической 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	
41.03.01 «Зарубежное регионоведение» (Б-ЗР)	ОПК-3 : Способен выделять, систематизировать и интерпретировать содержательно значимые эмпирические данные из потоков информации, а также смысловые конструкции в оригинальных текстах и источниках по профилю деятельности	ОПК-3.1к : Определяет содержательно значимые эмпирические данные из потоков информации	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
			РД2	Умение	применять широкий инструментарий статистического анализа данных при решении профессиональных задач, практической деятельности, а также при принятии управленческих решений
			РД3	Навык	проведения статистических расчетов с применением компьютерной техники и современных статистических пакетов прикладных программ

## 2 Место дисциплины (модуля) в структуре ОПОП

Дисциплина «Информационная безопасность и защита информации» 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 обучения	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо-емкость (З.Е.)	Объем контактной работы (час)					СРС	Форма аттес-тации	
					Всего	Аудиторная			Внеауди-торная			
						лек.	прак.	лаб.	ПА			КСР
41.03.01 Зарубежное регионоведение	ОФО	Б1.Б	5	2	37	18	18	0	1	0	35	3

### 4 Структура и содержание дисциплины (модуля)

#### 4.1 Структура дисциплины (модуля) для ОФО

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 ОФО

№	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	2	2	0	2	Тестовые задания, практические работы
2	Статистический анализ выборочных данных	РД1, РД2, РД3	4	0	4	4	Тестовые задания, практические работы
3	Проверка статистических гипотез	РД1, РД2, РД3	4	0	4	4	Тестовые задания, практические работы
4	Сравнение и создание групп	РД1, РД2, РД3	2	0	2	4	Тестовые задания, практические работы
5	Исследование статистических взаимосвязей	РД1, РД2, РД3	4	0	4	4	Тестовые задания, практические работы
6	Актуальность и роль визуализации данных в аналитике больших данных	РД1, РД2, РД3	2	0	2	2	Тестовые задания, практические работы
<b>Итого по таблице</b>			<b>18</b>	<b>2</b>	<b>16</b>	<b>20</b>	

#### 4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

*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: 19.11.2024). — 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: 14.11.2024). — 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: 14.11.2024). — 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: 30.09.2024)

**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.prlib.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 OfficeProfessionalPlus 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)

41.03.01 Зарубежное регионоведение. Политическая, правовая и деловая культура стран  
Азии

Год набора на ОПОП  
2024

Форма обучения  
очная

Владивосток 2024



## 1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
41.03.01 «Зарубежное регионоведение» (Б-ЗР)	ОПК-3 : Способен выделять, систематизировать и интерпретировать содержательно значимые эмпирические данные из потоков информации, а также смысловые конструкции в оригинальных текстах и источниках по профилю деятельности	ОПК-3.1к : Определяет содержательно значимые эмпирические данные из потоков информации

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

**Компетенция ОПК-3 «Способен выделять, систематизировать и интерпретировать содержательно значимые эмпирические данные из потоков информации, а также смысловые конструкции в оригинальных текстах и источниках по профилю деятельности»**

Таблица 2.1 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код результата	Тип результата	Результат	
ОПК-3.1к : Определяет содержательно значимые эмпирические данные из потоков информации	РД1	Знание	теоретические аспекты статистического исследования различных социально-экономических процессов на макро- и микроуровне, на основе познания конкретных методов и методов расчета и анализа экономических показателей	тест
	РД2	Умение	применять широкий инструментарий статистического анализа данных при решении профессиональных задач, практической деятельности, а также при принятии управленческих решений	решение задач
	РД3	Навык	проведения статистических расчетов с применением компьютерной техники и современных статистических пакетов прикладных программ	решение задач

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	
Очная 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 : применять широкий инструментарий статистического анализа данных при решении профессиональных задач, практической деятельности, а также при принятии управленческих решений	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
РД3	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, обнаруживает 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 и 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) частостей; 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

Цена (руб)	Количество проданных единиц
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**

### **Часть 1. Теоретические основы статистического анализа данных**

- 1. Что измеряет стандартное отклонение?**
  - а) Центральную тенденцию данных
  - б) Разброс значений относительно среднего
  - в) Максимальное значение в выборке
- 2. Какое из следующих утверждений верно?**
  - а) Среднее значение устойчиво к выбросам
  - б) Медиана менее чувствительна к выбросам, чем среднее
  - в) Размах показывает типичное значение выборки
- 3. Какой из следующих критериев используется для проверки гипотезы о равенстве средних значений двух выборок?**
  - а) t-критерий Стьюдента
  - б)  $\chi^2$ -критерий
  - в) Критерий Фишера
- 4. Что такое p-уровень значимости?**
  - а) Вероятность отклонения нулевой гипотезы
  - б) Вероятность случайного совпадения результатов
  - в) Вероятность получения наблюдаемых данных при условии верности нулевой гипотезы
- 5. Что показывает коэффициент корреляции Пирсона?**
  - а) Причинно-следственную связь между переменными
  - б) Линейную зависимость между двумя переменными
  - в) Степень вариативности одной переменной относительно другой
- 6. Чему равен коэффициент корреляции, если переменные полностью независимы?**
  - а) 0
  - б) 1
  - в) -1
- 7. Какой тип регрессионного анализа используется для прогнозирования категориальной переменной?**
  - а) Линейная регрессия
  - б) Логистическая регрессия
  - в) Множественная регрессия
- 8. Что называют генеральной совокупностью?**
  - а) Все элементы выборки
  - б) Весь массив данных, который представляет объект исследования
  - в) Набор переменных в выборке

9. Какой метод используют для уменьшения влияния выбросов на результаты анализа?

- a) Использование стандартного отклонения
- b) Применение медианы вместо среднего
- c) Применение логарифмирования

10. Что такое уровень значимости  $\alpha$ ?

- a) Максимально допустимая вероятность ошибочного принятия нулевой гипотезы
- b) Разница между средней выборки и генеральной совокупности
- c) Вероятность успеха эксперимента

1	2	3	4	5
b	b	a	c	b
6	7	8	9	10
a	b	b	b	a

## Часть 2. Практика: статистический анализ данных

1. 1. В выборке из 10 значений: 2, 4, 4, 4, 5, 5, 7, 9. Чему равна мода?

- a) 4
- b) 5
- c) Моды нет

2. Если стандартное отклонение равно 5, чему равна дисперсия?

- a) 16
- b) 25
- c) 0

3. Для теста гипотезы  $t = 2.5$ , критическое значение  $t = 2.3$ . Что следует сделать с нулевой гипотезой?

- a) принять
- b) отклонить

4. Если коэффициент корреляции Пирсона равен -0.8, что это означает?

- a) Сильная отрицательная линейная зависимость между переменными
- b) Умеренная положительная связь между переменными
- c) Связи нет

5. Чему равна сумма отклонений всех наблюдений от среднего значения?

- a) 0
- b) 100
- c) -1

6. В линейной регрессии коэффициент наклона ( $\beta_1$ ) равен 2. Что это означает?

- a) При увеличении независимой переменной на 2 единицы зависимая переменная увеличивается на 1 единицу.
- b) При увеличении независимой переменной на 1 единицу зависимая переменная увеличивается на 2 единицы.
- c) Сильная прямая зависимость между переменными

7. Если размер выборки увеличивается, то стандартная ошибка среднего:

- a) Увеличивается
- b) Уменьшается
- c) Остается неизменной

8. Какой из следующих инструментов лучше всего подходит для визуализации взаимосвязи между двумя переменными?

- a) Гистограмма
- b) Диаграмма рассеяния
- c) Коробчатая диаграмма

9. Что следует делать, если распределение выборки сильно скошено?

- a) Преобразовать данные (например, логарифмирование) или использовать методы непараметрической статистики.



- b) Удалить наблюдения
- c) Заменить на среднее

**10. Для проверки гипотезы о независимости двух категориальных переменных используют:**

- a) t-критерий
- b)  $\chi^2$ -тест
- c) ANOVA

Ключи к тесту «Практика: статистический анализ данных»				
1	2	3	4	5
a	b	b	a	a
6	7	8	9	10
b	b	b	a	b

### Часть 3. Современные методы визуализации данных

1. **Что такое Power BI?**
2. Инструмент для работы с базами данных.
3. Платформа для визуализации данных и построения отчетов.
4. Средство для создания презентаций.
  1. **Какой тип соединения позволяет работать с большими объемами данных, загружая их по мере необходимости?**
  2. Импорт данных.
  3. Прямое подключение (DirectQuery).
  4. Хранение в кэше.
    1. **Какие источники данных можно подключить к Power BI?**
    2. Только Excel и базы данных SQL.
    3. Только веб-сервисы.
    4. Различные источники, включая облачные, локальные и веб-ресурсы.
      1. **Какой элемент Power BI используется для настройки взаимодействия между диаграммами?**
      2. Фильтры.
      3. Модель данных.
      4. Срезы.
        1. **Что такое дашборд в Power BI?**
        2. Набор статических отчетов.
        3. Интерактивная панель, объединяющая несколько визуализаций.
        4. Список подключенных источников данных.
          1. **Какая функция Power BI используется для трансформации данных?**
          2. Power Query.
          3. Power Editor.
          4. Data Transformer.
            1. **Какой тип визуализации лучше всего подходит для отображения временных рядов?**
            2. Круговая диаграмма.
            3. Линейный график.
            4. Столбчатая диаграмма.
              1. **Что позволяет делать функция Drillthrough?**
              2. Применять фильтры ко всем отчетам.
              3. Просматривать детализированную информацию на отдельной странице.
              4. Объединять несколько визуализаций.
                1. **Как настроить автоматическое**

## **обновление данных в дашборде Power BI?**

2. Использовать DirectQuery.
3. Настроить расписание обновлений.
4. Экспортировать отчет в Excel.

## **10. Какие языки можно использовать для написания вычисляемых столбцов и мер в Power BI?**

1. SQL и Python.
2. DAX и R.
3. Только SQL.

## **11. Что из перечисленного является основной функцией Data Lens?**

1. Хранение данных.
2. Визуализация и аналитика больших данных.
3. Создание баз данных.

## **12. Какой тип данных наиболее эффективно обрабатывается Data Lens?**

1. Только структурированные данные.
2. Неструктурированные данные в реальном времени.
3. Как структурированные, так и неструктурированные данные.

## **13. Какой тип карты можно построить в Data Lens?**

1. Тематическую карту с геоданными.
2. Только точечную карту.
3. Карту, ограниченную региональными данными.

## **14. Какая функция Data Lens используется для отображения данных по временным меткам?**

1. Временная шкала.
2. График событий.
3. Временной ряд.

## **15. Что такое потоковая обработка данных в Data Lens?**

1. Одновременная обработка и визуализация входящих данных.
2. Пакетная обработка больших объемов данных.
3. Обработка данных через облачное хранилище.

## **16. Какая настройка позволяет сделать отчет интерактивным?**

1. Добавление слоев данных.
2. Использование фильтров.
3. Экспорт отчета в PDF.

## **17. Какая визуализация наиболее подходит для отображения плотности данных в регионе?**

1. Линейный график.
2. Тепловая карта.
3. Круговая диаграмма.

## **18. Какой элемент интерфейса Data Lens используется для добавления новых источников данных?**

1. Панель управления.
2. Настройка рабочего пространства.
3. Менеджер подключений.

**19. Какая функция позволяет объединить данные из нескольких таблиц?**

1. Слияние данных.
2. Связывание таблиц.
3. Группировка данных.

**20. Какой формат отчета поддерживает Data Lens для публикации?**

1. Только PDF.
2. Только интерактивный веб-формат.
3. Как статические, так и интерактивные форматы.

Ключи к тесту «Современные методы визуализации данных»				
1	2	3	4	5
b	b	c	c	b
6	7	8	9	10
a	b	b	b	b
11	12	13	14	15
b	c	a	c	a
16	17	18	19	20
b	b	c	a	c

*Краткие методические указания*

Тест содержит задания с выбором одного или нескольких ответов.

*Шкала оценки*

Оценка	Баллы	Описание
5	19-20	Тесты решены правильно.
4	16–18	Допущено не более 2-х ошибок
3	9-16	Допущено не более 4-х ошибок
2	1-9	Допущено более 5-х ошибок