

*Федеральное государственное автономное образовательное учреждение высшего образования  
«Российский университет дружбы народов»*

*Факультет физико-математических и естественных наук*

Рекомендовано МССН  
38.00.00 «Экономика и управление»,  
подгруппа 4 «Бизнес-информатика»

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Наименование дисциплины**

Анализ данных

---

**Рекомендуется для направления подготовки/специальности**

38.03.05 Бизнес-информатика

---

**Квалификация (степень) выпускника**

бакалавр

## 1. Цели и задачи дисциплины:

Целью дисциплины является знакомство слушателей с основными методами и моделями статистического анализа данных для обработки массивов информации различного рода, изучение способов применения таких методов и анализа их эффективности в конкретных случаях.

В рамках дисциплины решаются следующие задачи: развитие у слушателей навыков формализации статистической задачи, изучение математической основы задачи анализа данных, применение изучаемых методов для решения конкретных задач анализа, формирование умения строить верные выводы на основе результатов анализа.

## 2. Место дисциплины в структуре ОП ВО:

Дисциплина Анализ данных относится к обязательной части блока Б1 учебного плана.

В таблице № 1 приведены предшествующие и последующие дисциплины, направленные на формирование компетенций дисциплины в соответствии с матрицей компетенций ОП ВО.

Таблица № 1

### Предшествующие и последующие дисциплины, направленные на формирование компетенций

№ п/п	Шифр и наименование компетенции	Предшествующие дисциплины	Последующие дисциплины (группы дисциплин)
1.	УК-1, УК-12	Архитектура предприятия; Основы разработки корпоративных инфокоммуникационных систем	Статистический анализ Интеллектуальный анализ данных
<b>Общепрофессиональные компетенции</b>			
2.	ОПК-1, ОПК-2	Архитектура предприятия; Основы разработки корпоративных инфокоммуникационных систем; Системы поддержки принятия решений	Модуль «Математическое моделирование в бизнес-информатике» Методы машинного обучения; Интеллектуальный анализ данных
3.	ОПК-4, ОПК-6	Основы разработки корпоративных инфокоммуникационных систем; Системы поддержки принятия решений	модуль «Математическое моделирование в бизнес-информатике»; Методы машинного обучения, Интеллектуальный анализ данных
4.	ОПК-7	Архитектура предприятия; Системы поддержки принятия решений	Статистический анализ модуль «Математическое моделирование в бизнес-информатике»; Методы машинного обучения
<b>Профессиональные компетенции (вид профессиональной деятельности <u>организационно-управленческая</u>)</b>			
5.	ПК-3	Основы разработки корпоративных инфокоммуникационных систем; Системы поддержки принятия решений	Преддипломная практика Защита ВКР
<b>Профессиональные компетенции (вид профессиональной деятельности <u>аналитическая</u>)</b>			

6.	ПК-4	Системы поддержки принятия решений	Математическое моделирование в бизнес-информатике Методы машинного обучения Интеллектуальный анализ данных
----	------	------------------------------------	--

### 3. Требования к результатам освоения дисциплины:

Процесс изучения дисциплины направлен на формирование следующих компетенций: УК-1, УК-12, ОПК-1, ОПК-4, ОПК-6, ОПК-7, ПК-3, ПК-4.

УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

- УК-1.2 Умеет анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности
- УК-1.3 Владеет навыками научного поиска и практической работы с информационными источниками; методами принятия решений

УК-12 Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных

УК-12.1 Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных

ОПК-1 Способен проводить моделирование, анализ и совершенствование бизнес-процессов и информационно-технологической инфраструктуры предприятия в интересах достижения его стратегических целей с использованием современных методов и программного инструментария

- ОПК-1.1 Знает инструменты и методы моделирования бизнес-процессов
- ОПК-1.2 Знает методы анализа ИТ-инфраструктуры предприятия
- ОПК-1.3 Умеет проводить анализ ИТ-инфраструктуры предприятия

ОПК-2 Способен проводить исследование и анализ рынка информационных систем и информационно-коммуникационных технологий, выбирать рациональные решения для управления бизнесом

- ОПК-2.1 Знает методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации для проведения исследования рынка ИС и ИКТ
- ОПК-2.2 Умеет проводить сбор, анализ, систематизацию информации для проведения исследования рынка ИС и ИКТ

ОПК-3 Способен управлять процессами создания и использования продуктов и услуг в сфере информационно-коммуникационных технологий, в том числе разрабатывать алгоритмы и программы для их практической реализации

- ОПК-3.1 Знает современные инструменты и методы управления процессами разработки и применения продуктов и услуг в сфере ИКТ
- ОПК-3.2 Знает современные стандарты информационного взаимодействия систем

ОПК-4 Способен понимать принципы работы информационных технологий; использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений

- ОПК-4.1 Знает методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации для проведения бизнес-анализа
- ОПК-4.2 Умеет применять информационные технологии в объеме, необходимом для бизнес-анализа
- ОПК-4.3 Умеет оформлять результаты бизнес-анализа в соответствии с выбранными подходами

ОПК-6 Способен выполнять отдельные задачи в рамках коллективной научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно коммуникационных технологий

ОПК-6.1 Способен выполнять отдельные задачи в рамках коллективной научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий

ОПК-7 Способен использовать цифровые технологии и методы в профессиональной деятельности в области бизнес-информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.

- ОПК-7.1 Знает базовые принципы цифровых технологий и методов, необходимых в профессиональной деятельности в области бизнес-информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.
- ОПК-7.2 Умеет применять необходимые в профессиональной деятельности цифровые технологии и методы в области бизнес-информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.
- ОПК-7.3 Владеет необходимыми в профессиональной деятельности технологиями и методами в области бизнес-информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.

ПК-3 Способен выполнять работы и управлять работами по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы

- ПК-3.1 Знает основы архитектуры, устройства и функционирования информационно-вычислительных систем и сетевых подсистем инфокоммуникационной системы организации; основы современных операционных систем; сетевые протоколы
- ПК-3.2 Знает теорию баз данных и основы современных систем управления базами данных
- ПК-3.3 Знает основы программирования; современные объектно-ориентированные языки программирования; современные структурные языки программирования; языки современных бизнес-приложений
- ПК-3.4 Умеет кодировать на языках программирования; тестировать результаты кодирования
- ПК-3.5 Умеет работать с базами данных, с системами управления базами данных

ПК-4: способность принимать обоснованные управленческие решения в своей профессиональной деятельности.

ПК-4 Способен принимать обоснованные управленческие решения в своей профессиональной деятельности

ПК-4.1 Знает языки визуального моделирования

ПК-4.2 Умеет анализировать и оценивать факторы и условия, влияющие на принятие управленческих решений

ПК-4.3 Умеет проводить оценку эффективности принятия решения в соответствии с выбранными критериями или выбранными целевыми показателями

В результате изучения дисциплины студент должен:

**Знать:**

- Методы построения точечных оценок и доверительных интервалов параметров случайных величин;
- Общий алгоритм проверки статистических гипотез;
- Параметры статических критериев – уровень значимости, доверительная вероятность, мощность критерия;
- Методы проверки однородности выборочных данных;
- Методы проверки независимости признаков;
- Коэффициенты, измеряющие силу связи показателей;
- Способы проведения анализа с использованием программных средств.

**Уметь:**

- Применять в профессиональной, исследовательской и прикладной деятельности статистические критерии для проверки гипотез относительно изучаемых данных;
- Оценивать неизвестные параметры анализируемых данных;
- Осуществлять обработку статистической информации в условиях стохастической неопределенности;
- Проводить сравнительный анализ различных статистических критериев при исследовании конкретного набора данных;
- Использовать программные средства (статистические пакеты, языки программирования) для проведения анализа данных.

**Владеть:**

- Основными определениями, методами и моделями статистического анализа данных;
- Способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям;
- Навыком использования программных средств для проведения анализа данных.

**4. Объем дисциплины и виды учебной работы**

Общая трудоемкость дисциплины составляет 4 зачетных единиц.

Вид учебной работы	Всего часов	Семестры
		Семестр 6., мод. 11
<b>Аудиторные занятия (всего)</b>	36	36
<i>Лекции</i>	18	18
<i>Практические занятия (ПЗ)</i>		
<i>Семинары (С)</i>		
<i>Лабораторные работы (ЛР)</i>	18	18
<b>Самостоятельная работа (всего)</b>	81	81
<b>Контроль знаний</b>	27	27
<b>Общая трудоемкость (часов)</b>	144	144
(зач. ед.)	4	4

## 5. Содержание дисциплины

### 5.1. Содержание разделов дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела
1.	Выборочные исследования	1. Простой случайный отбор, репрезентативная выборка. Выбор без возвращения. 2. Точечные оценки неизвестных параметров. Оценивание долей и процентов. Свойства оценок. 3. Доверительные интервалы для неизвестных параметров случайных величин.
2.	Статистические гипотезы. Исследование однородности	1. Проверка статистических гипотез. Параметры критериев. Уровень значимости, доверительная вероятность, мощность критерия. Ошибки I и II рода. Ранг и вариационный ряд. 2. Однородности выборок. Неоднородность вида смещение. Классический критерий Стьюдента, критерий Вилкоксона. 3. Однородности выборок. Неоднородность вида сжатие/растяжение. Критерий Фишера.
3.	Анализ статистической взаимосвязи	1. Независимость случайных величин. Исследование зависимости между количественными и порядковыми признаками. 2. Выборочный коэффициент корреляции. Коэффициент корреляции Спирмена. Коэффициент согласованности Кендалла. 3. Исследование зависимости между номинальными признаками. Критерий хи-квадрат.

### 5.2. Разделы дисциплин и виды занятий

№ п/п	Наименование раздела дисциплины	Лекц.	Практ. зан.	Лаб. зан.	Семина	СРС, контроль	Всего час.
1.	Выборочные исследования	4				24	28
2.	Статистические гипотезы. Исследование однородности	8		6		36	50
3.	Анализ статистической взаимосвязи	6		12		48	66
	<b>Итого</b>	18		18		108	144

## 6. Лабораторный практикум

№ п/п	№ раздела дисциплины	Темы лабораторных работ	Трудо-емкость (час.)
1.	2	Применение пакетов статистического анализа для проверки гипотез об однородности вида смещение в двухвыборочной модели	2
2.	2	Применение пакетов статистического анализа для проверки гипотез об однородности в двухвыборочной модели (в общем случае)	4
3.	3	Применение пакетов статистического анализа для	6

		исследования зависимости между количественными признаками	
4.	3	Применение пакетов статистического анализа для исследования зависимости между количественными и номинальными признаками	6

## 7. Практические занятия (семинары) не предусмотрены.

## 8. Материально-техническое обеспечение дисциплины:

Учебная аудитория с ПК и проектором для проведения учебных занятий (в том числе для практического и лекционного типов занятий, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации).

Компьютерные (дисплейные) классы с доступом к сети Интернет и электронно-образовательной среде Университета для выполнения обучающимися лабораторного практикума, самостоятельной работы и проведения компьютерного тестирования обучающихся (при необходимости).

## 9. Информационное обеспечение дисциплины:

а) программное обеспечение:

продукты Microsoft - операционная система, пакет офисных приложений, MS Teams и др. (подписка Enrollment for Education Solutions (EES)); программное обеспечение со свободной лицензией: операционная система Linux (дистрибутив Gentoo), LibreOffice, SciLab, Python.

б) базы данных, информационно-справочные и поисковые системы

- Сайт библиотеки РУДН <http://lib.rudn.ru/>.
- Сайт ТУИС <http://esystem.pfur.ru/>.

## 10. Учебно-методическое обеспечение дисциплины:

а) основная литература:

1. Горяинова, Е.Р. Прикладные методы анализа статистических данных : учебное пособие / Е.Р. Горяинова, А.Р. Панков, Е.Н. Платонов. - Москва : Издательский дом Высшей школы экономики, 2012. - 312 с. - ISBN 978-5-7598-0866-4 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=227280>

б) дополнительная литература:

1. Таблицы математической статистики [Текст] / Л.Н. Большев, Н.В. Смирнов. - М. : Наука, 1965. - 464 с. - 3.50.
2. Крутиков, В.Н. Анализ данных : учебное пособие / В.Н. Крутиков, В.В. Мешечкин ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово : Кемеровский государственный университет, 2014. - 138 с. : ил. - Библиогр. в кн. - ISBN 978-5-8353-1770-7 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=278426>
2. Лисицин, Д.В. Устойчивые методы оценивания параметров статистических моделей : учебное пособие / Д.В. Лисицин. - Новосибирск : НГТУ, 2013. - 76 с. - ISBN 978-5-7782-2196-3 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=228831>

## 11. Методические указания для обучающихся по освоению дисциплины (модуля)

Учебным планом на изучение дисциплины отводится один семестр. В течение семестра проводятся лекции, выполняются лабораторные работы, контрольные мероприятия. В конце семестра проводится итоговый контроль знаний.

### **11.1. Методические указания по самостоятельному освоению теоретического материала по дисциплине**

Лекционный материал дисциплины охватывает темы, указанные в разделе 5.1 программы дисциплины. В ТУИС (<http://esystem.pfur.ru>) по темам лекций размещены дополнительные материалы. Также рекомендуется изучать конкретные разделы литературы, указанные в рамках каждой темы курса (полный список литературы приведен в п.10 программы).

В ТУИС также размещены прикладные задачи, относящиеся к каждому разделу курса. Рекомендуется проработка методов решения таких задач с использованием программных пакетов. Подробное описание можно получить в литературе, указанной в п.10 программы.

### **11.2. Методические указания по выполнению лабораторных работ**

Задания по лабораторным работам выполняются индивидуально каждым студентом в дисплейных классах в соответствии с календарным планом и методическими указаниями по выполнению лабораторных работ по дисциплине.

Лабораторные работы направлены на формирование практических навыков у студентов по постановке и решению различных задач анализа данных с использованием статистических пакетов.

По результатам выполнения каждой лабораторной работы студентом готовится отчет. Отчеты в электронном виде сдаются студентом на проверку через соответствующий раздел ТУИС (<http://esystem.pfur.ru>).

### **11.3. Методические указания по подготовке к контрольным мероприятиям**

Контрольные мероприятия по дисциплине проводятся в форме контрольных работ, онлайн-тестирований и оценки результатов выполнения лабораторных работ. Задания (темы) для подготовки к выполнению контрольных работ и прохождению тестирований размещены в соответствующем разделе ТУИС (<http://esystem.pfur.ru>).

Итоговый контроль в форме опроса проводится по темам всех разделов дисциплины. Вопросы для подготовки к промежуточному и итоговому контролю размещены в соответствующем разделе ТУИС (<http://esystem.pfur.ru>).

## **12. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине (модулю)**

ФОС по дисциплине представлен в приложении к данной программе.

Программа составлена в соответствии с требованиями ОС ВО РУДН.

**Разработчик:**

**Зав. кафедрой** информационных технологий, д.ф.-м.н.



Ю.Н. Орлов

**Руководитель программы**

заведующий кафедрой прикладной информатики и теории вероятностей, д.т.н., проф.



К.Е. Самуилов



Федеральное государственное автономное образовательное учреждение высшего образования  
«Российский университет дружбы народов»

Факультет физико-математических и естественных наук

**Кафедра информационных технологий**

# **ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**

*ПО УЧЕБНОЙ ДИСЦИПЛИНЕ*

---

*Анализ данных*

(наименование дисциплины)

**38.03.05 — Бизнес-информатика**

(код и наименование направления подготовки)

---

(наименование профиля подготовки)

**бакалавр**

---

Квалификация (степень) выпускника

## Паспорт фонда оценочных средств по дисциплине Анализ данных

Направление: 38.03.05 — Бизнес-информатика

Код контролируемой компетенции или ее части	Контролируемый раздел дисциплины	Контролируемая тема дисциплины	ФОСы (формы контроля уровня освоения ООП)				Баллы темы	Баллы раздела
			Аудиторная работа			Экзамен/Зачет		
			Контрольная работа	Выполнение ЛР	Тест			
УК-1, УК-12, ОПК-1, ОПК-4, ОПК-6, ОПК-7, ПК-3, ПК-4.	Раздел 1: Выборочные исследования	Тема 1: Простой случайный отбор, репрезентативная выборка. Выбор без возвращения.			1	3	4	28
		Тема 2: Точечные оценки неизвестных параметров. Оценивание долей и процентов. Свойства оценок.	5		2	3	10	
		Тема 3: Доверительные интервалы для неизвестных параметров случайных величин	10		2	2	14	
	Раздел 2: Статистические гипотезы. Исследование однородности	Тема 1: Проверка статистических гипотез. Параметры критериев. Уровень значимости, доверительная вероятность, мощность критерия. Ошибки I и II рода. Ранг и вариационный ряд.			5	2	7	37
		Тема 2: Однородности выборок. Неоднородность вида смещение. Классический критерий Стьюдента, критерий Вилкоксона.	5	8		2	15	
		Тема 3: Однородности выборок. Неоднородность вида сжатие/растяжение. Критерий Фишера.	5	8		2	15	
	Раздел 3: Анализ статистической взаимосвязи	Тема 1: Независимость случайных величин. Исследование зависимости между количественными и порядковыми признаками.			5	2	7	35
		Тема 2: Выборочный коэффициент корреляции. Коэффициент корреляции Спирмена. Коэффициент согласованности Кендалла.		12		2	14	
		Тема 3: Исследование зависимости между номинальными признаками. Критерий хи-квадрат.		12		2	14	
<b>ИТОГО:</b>			25	40	15	20	100	100

Процесс изучения дисциплины направлен на формирование следующих компетенций УК-1, УК-12, ОПК-1, ОПК-4, ОПК-6, ОПК-7, ПК-3, ПК-4.

*(указываются в соответствии с ОС ВО РУДН)*

- УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач
- УК-1.2 Умеет анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности
- УК-1.3 Владеет навыками научного поиска и практической работы с информационными источниками; методами принятия решений
- УК-12 Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных
- УК-12.1 Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных
- ОПК-1 Способен проводить моделирование, анализ и совершенствование бизнес-процессов и информационно-технологической инфраструктуры предприятия в интересах достижения его стратегических целей с использованием современных методов и программного инструментария
- ОПК-1.1 Знает инструменты и методы моделирования бизнес-процессов
- ОПК-1.2 Знает методы анализа ИТ-инфраструктуры предприятия
- ОПК-1.3 Умеет проводить анализ ИТ-инфраструктуры предприятия
- ОПК-2 Способен проводить исследование и анализ рынка информационных систем и информационно-коммуникационных технологий, выбирать рациональные решения для управления бизнесом
- ОПК-2.1 Знает методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации для проведения исследования рынка ИС и ИКТ
- ОПК-2.2 Умеет проводить сбор, анализ, систематизацию информации для проведения исследования рынка ИС и ИКТ
- ОПК-3 Способен управлять процессами создания и использования продуктов и услуг в сфере информационно-коммуникационных технологий, в том числе разрабатывать алгоритмы и программы для их практической реализации
- ОПК-3.1 Знает современные инструменты и методы управления процессами разработки и применения продуктов и услуг в сфере ИКТ
- ОПК-3.2 Знает современные стандарты информационного взаимодействия систем
- ОПК-4 Способен понимать принципы работы информационных технологий; использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений
- ОПК-4.1 Знает методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации для проведения бизнес-анализа
- ОПК-4.2 Умеет применять информационные технологии в объеме, необходимом для бизнес-анализа
- ОПК-4.3 Умеет оформлять результаты бизнес-анализа в соответствии с выбранными подходами

ОПК-6 Способен выполнять отдельные задачи в рамках коллективной научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно коммуникационных технологий

- ОПК-6.1 Способен выполнять отдельные задачи в рамках коллективной научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий

ОПК-7 Способен использовать цифровые технологии и методы в профессиональной деятельности в области бизнес-информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.

- ОПК-7.1 Знает базовые принципы цифровых технологий и методов, необходимых в профессиональной деятельности в области бизнес-информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.
- ОПК-7.2 Умеет применять необходимые в профессиональной деятельности цифровые технологии и методы в области бизнес-информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.
- ОПК-7.3 Владеет необходимыми в профессиональной деятельности технологиями и методами в области бизнес-информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.

ПК-3 Способен выполнять работы и управлять работами по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы

- ПК-3.1 Знает основы архитектуры, устройства и функционирования информационно-вычислительных систем и сетевых подсистем инфокоммуникационной системы организации; основы современных операционных систем; сетевые протоколы
- ПК-3.2 Знает теорию баз данных и основы современных систем управления базами данных
- ПК-3.3 Знает основы программирования; современные объектно-ориентированные языки программирования; современные структурные языки программирования; языки современных бизнес-приложений
- ПК-3.4 Умеет кодировать на языках программирования; тестировать результаты кодирования
- ПК-3.5 Умеет работать с базами данных, с системами управления базами данных

ПК-4: способность принимать обоснованные управленческие решения в своей профессиональной деятельности.

ПК-4 Способен принимать обоснованные управленческие решения в своей профессиональной деятельности

- ПК-4.1 Знает языки визуального моделирования
- ПК-4.2 Умеет анализировать и оценивать факторы и условия, влияющие на принятие управленческих решений
- ПК-4.3 Умеет проводить оценку эффективности принятия решения в соответствии с выбранными критериями или выбранными целевыми показателями

# Балльно-рейтинговая система оценки уровня знаний

Таблица соответствия баллов и оценок

Баллы БРС	Традиционные оценки РФ	Оценки ECTS
95 - 100	5	A
86 - 94		B
69 - 85	4	C
61 - 68	3	D
51 - 60		E
31 - 50	2	FX
0 - 30		F
51-100	Зачет	Passed

## Правила применения БРС

1. Раздел (тема) учебной дисциплины считаются освоенными, если студент набрал более 50 % от возможного числа баллов по этому разделу (теме).
2. Студент не может быть аттестован по дисциплине, если он не освоил все темы и разделы дисциплины, указанные в сводной оценочной таблице дисциплины.
3. По решению преподавателя и с согласия студентов, не освоивших отдельные разделы (темы) изучаемой дисциплины, в течение учебного семестра могут быть повторно проведены мероприятия текущего контроля успеваемости или выданы дополнительные учебные задания по этим темам или разделам. При этом студентам за данную работу засчитывается минимально возможный положительный балл (51 % от максимального балла).
4. При выполнении студентом дополнительных учебных заданий или повторного прохождения мероприятий текущего контроля полученные им баллы засчитываются за конкретные темы. Итоговая сумма баллов не может превышать максимального количества баллов, установленного по данным темам (в соответствии с приказом Ректора № 564 от 20.06.2013). По решению преподавателя предыдущие баллы, полученные студентом по учебным заданиям, могут быть аннулированы.
5. График проведения мероприятий текущего контроля успеваемости формируется в соответствии с календарным планом курса. Студенты обязаны сдавать все задания в сроки, установленные преподавателем.
6. Время, которое отводится студенту на выполнение мероприятий текущего контроля успеваемости, устанавливается преподавателем. По завершение отведенного времени студент должен сдать работу преподавателю, вне зависимости от того, завершена она или нет.
7. Использование источников (в том числе конспектов лекций и лабораторных работ) во время выполнения контрольных мероприятий возможно только с разрешения преподавателя.
8. Отсрочка в прохождении мероприятий текущего контроля успеваемости считается уважительной только в случае болезни студента, что подтверждается наличием у него медицинской справки, заверенной круглой печатью в поликлинике № 25,

предоставляемой преподавателю не позднее двух недель после выздоровления. В этом случае выполнение контрольных мероприятий осуществляется после выздоровления студента в срок, назначенный преподавателем. В противном случае, отсутствие студента на контрольном мероприятии признается не уважительным.

9. Студент допускается к итоговому контролю знаний с любым количеством баллов, набранных в семестре.
10. Итоговая контроль знаний оценивается из 20 баллов независимо от числа баллов за семестр.
11. Если в итоге за семестр студент получил менее 31 балла, то ему выставляется оценка F и студент должен повторить эту дисциплину в установленном порядке. Если же в итоге студент получил 31-50 баллов, т. е. FX, то студенту разрешается добор необходимого (до 51) количества баллов путем повторного однократного выполнения предусмотренных контрольных мероприятий, при этом по усмотрению преподавателя аннулируются соответствующие предыдущие результаты. Ликвидация задолженностей проводится в период с 07.02 по 28.02 (с 07.09 по 28.09) по согласованию с деканатом.

## Примерный перечень оценочных средств

п/п	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде
<i>Аудиторная работа</i>			
1	Контрольная работа	Средство контроля, организованное как аудиторное занятие, на котором обучающимся необходимо самостоятельно продемонстрировать усвоение учебного материала темы, раздела или разделов дисциплины.	Вопросы по темам/разделам дисциплины
2	Лабораторная работа	Система практических заданий, направленных на формирование практических навыков у обучающихся	Фонд практических заданий
3	Тест	Система стандартизированных заданий (вопросов), позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося.	База тестовых заданий
4	Экзамен	Оценка работы студента в течение семестра (года, всего срока обучения и др.) и призван выявить уровень, прочность и систематичность полученных им теоретических и практических знаний, приобретения навыков самостоятельной работы, развития творческого мышления, умение синтезировать полученные знания и применять их в решении практических задач.	Примеры заданий/вопросов, пример экзаменационного билета

Учебным планом на изучение дисциплины отводится один семестр. В дисциплине предусмотрены лекции, контрольная работа, лабораторный практикум, тестирования. В конце семестра проводится итоговый контроль знаний.

Оценивание результатов освоения дисциплины производится в соответствии с балльно-рейтинговой системой. По дисциплине предусмотрен экзамен.

Итоговый контроль знаний по дисциплине проводится в форме опроса. При необходимости экзамен может проводиться в форме письменного ответа на вопросы из билетов.

## Критерии оценки по дисциплине

Итоговая оценка выставляется по сумме набранных баллов за контрольную работу, четыре лабораторные работы и три тестирования. На промежуточную аттестацию отводится 20 баллов, и она является обязательной для всех студентов.

*95-100 баллов:*

- полное выполнение контрольной работы, лабораторных работ и прохождения тестирований;
- решение всех заданий промежуточной аттестации;
- активное участие в мероприятиях, предусмотренных программой дисциплины;
- систематизированное, глубокое и полное освоение навыков и компетенций по всем разделам программы дисциплины;
- использование научной терминологии, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;
- умение эффективно использовать методику программы дисциплины в постановке и решении научных и профессиональных задач;
- выраженная способность самостоятельно и творчески решать поставленные задачи;
- полная самостоятельность и творческий подход при изложении материала по программе дисциплины;
- полное и глубокое усвоение основной и дополнительной литературы, рекомендованной программой дисциплины и преподавателем.

*86- 94 балла:*

- полное выполнение контрольной работы, лабораторных работ и прохождения тестирований;
- решение практических заданий промежуточной аттестации;
- участие в мероприятиях, предусмотренных программой дисциплины;
- систематизированное, глубокое и полное освоение навыков и компетенций по всем разделам программы дисциплины;
- использование научной терминологии, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;
- умение эффективно использовать методику программы дисциплины в постановке и решении научных и профессиональных задач;
- способность самостоятельно решать поставленные задачи в нестандартных производственных ситуациях;
- усвоение основной и дополнительной литературы, рекомендованной программой дисциплины и преподавателем.

*69-85 баллов:*

- частичное выполнение контрольной работы, лабораторных работ и прохождения тестирований;
- решение практических заданий промежуточной аттестации;
- участие в мероприятиях, предусмотренных программой дисциплины;
- систематизированное и полное освоение навыков и компетенций по всем разделам программы дисциплины;
- умение использовать методику программы дисциплины в постановке и решении научных и профессиональных задач;
- усвоение основной литературы, рекомендованной программой дисциплины.

*51-68 баллов:*



- частичное выполнение контрольных работ, лабораторных работ и прохождение тестирований;
- решение практических заданий промежуточной аттестации;
- участие в мероприятиях, предусмотренных программой дисциплины;
- систематизированное и полное освоение навыков и компетенций по всем разделам программы дисциплины;
- удовлетворительное умение использовать методику программы дисциплины в постановке и решении научных и профессиональных задач;
- удовлетворительное усвоение основной литературы;

*31 - 50 баллов – НЕ ЗАЧТЕНО:*

- частичное выполнение контрольных работ, либо частичное выполнение лабораторных работ;
- недостаточно полный объем навыков и компетенции в рамках программы дисциплины;
- неумение использовать в практической деятельности научной терминологии, изложение ответа на вопросы с существенными стилистическими и логическими ошибками;
- слабое умение использовать методику программы дисциплины в постановке и решении научных и профессиональных задач;
- удовлетворительное усвоение основной литературы.

*0-30 баллов, НЕ ЗАЧТЕНО:*

- отсутствие умений, навыков, знаний и компетенции в рамках программы дисциплины;
- невыполнение лабораторных заданий; невыполнение контрольных работ; отказ от ответа по программе дисциплины;
- игнорирование занятий по дисциплине по неуважительной причине.

# Комплект экзаменационных билетов

Дисциплина Анализ данных  
(наименование дисциплины)

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1

1. Перечислите условия применимости критерия Спирмена.
2. Когда достигаются максимальные значения мер Пирсона и Крамера? Приведите пример.
3. Понятие и основные характеристики простого случайного отбора.

**Составитель**

Заведующий кафедрой

Ю. Н. Орлов

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №2

1. Критерии проверки однородности вида смещение.
2. Ошибки I и II рода при проверке статистических гипотез.
3. Количественные, порядковые и номинальные данные – описание и различия.

**Составитель**

Заведующий кафедрой

Ю. Н. Орлов

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №3

1. Критерий Вилкоксона. Назначение, условия применения, статистика и ее распределение.
2. Репрезентативная выборка, стратификация данных.
3. Основные различия между критериями Кендалла/Спирмена и критерием Хи-квадрат.

**Составитель**

Заведующий кафедрой

Ю. Н. Орлов

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №4

1. Критерий Фишера. Назначение, условия применения, статистика и ее распределение.
2. Свойства точечных оценок неизвестных параметров.
3. Стандартная ошибка оценки среднего значения.

**Составитель**

Заведующий кафедрой

Ю. Н. Орлов

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №5

1. Критерий Пирсона. Назначение, условия применения, статистика и ее распределение.
2. Уровень значимости и доверительная вероятность.
3. Различия между мерами Пирсона и Крамера для критерия Хи-квадрат.

**Составитель**

Заведующий кафедрой

Ю. Н. Орлов

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №6

1. Критерий Спирмена. Назначение, условия применения, статистика и ее распределение.
2. Вариационный ряд, операция ранжирования.
3. Критические области, р-значения.

**Составитель**

Заведующий кафедрой

Ю. Н. Орлов

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №7

1. Критерий Кендалла. Назначение, условия применения, статистика и ее распределение.
2. Коэффициент корреляции, выборочный коэффициент корреляции, коэффициент согласованности Кендалла.
3. Неоднородности вида смещение и сжатие/растяжение.

**Составитель**

Заведующий кафедрой

Ю. Н. Орлов

## ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №8

1. Критерий Хи-квадрат. Назначение, условия применения, статистика и ее распределение
2. Зависимость и независимость случайных величин
3. Состоятельность и несмещенность точечных оценок

**Составитель**

Заведующий кафедрой

Ю. Н. Орлов

## **Примерный перечень вопросов для опроса в ходе итогового контроля знаний**

по дисциплине Анализ данных

1. Виды шкал данных – количественная, порядковая, номинальная
2. Описание и основные характеристики простого случайного отбора
3. Генеральная совокупность, выборка, репрезентативная выборка
4. Выбор с возвращением и без возвращения
5. Примеры точечных оценок, формулы расчетов
6. Свойства точечных оценок
7. Методы оценивания долей и процентов
8. Дисперсия выборки и стандартная ошибка среднего
9. Доверительные интервалы для неизвестных параметров
10. Понятие статистической гипотезы, нулевая и альтернативные гипотезы
11. Параметры статистических критериев
12. Уровень значимости, доверительная вероятность, мощность критерия
13. Ошибки I и II рода
14. Ранг и вариационный ряд
15. Понятие однородности выборок
16. Неоднородность вида смещение
17. Неоднородность вида сжатие/растяжение
18. Односторонние и двусторонние критерии
19. Критерий Стьюдента – условия применимости
20. Критерий Стьюдента – статистика и ее распределение
21. Критерий Вилкоксона – условия применимости
22. Критерий Вилкоксона – статистика и ее распределение
23. Нулевая и альтернативные гипотезы критериев Стьюдента и Вилкоксона
24. Критерий Фишера – условия применимости
25. Критерий Фишера – статистика и ее распределение
26. Нулевая и альтернативные гипотезы критерия Фишера
27. Зависимость и независимость случайных величин по определению
28. Коэффициент корреляции, выборочный коэффициент корреляции
29. Критерий Пирсона – условия применимости
30. Критерий Пирсона – статистика и ее распределение
31. Нулевая и альтернативные гипотезы критерия Пирсона
32. Коэффициент согласованности Кендалла
33. Критерий Кендалла – условия применимости
34. Критерий Кендалла – статистика и ее распределение
35. Критерий Спирмена – условия применимости
36. Критерий Спирмена – статистика и ее распределение
37. Нулевая и альтернативные гипотезы критериев Кендалла и Спирмена
38. Таблица сопряженности
39. Критерий Хи-квадрат – условия применимости
40. Критерий Хи-квадрат – статистика и ее распределение
41. Нулевая и альтернативные гипотезы критерия Хи-квадрат
42. Меры Пирсона и Крамера

## Комплект заданий для контрольной работы

по дисциплине Анализ данных

**Тема** Точечные оценки неизвестных параметров генеральной совокупности. Построение доверительных интервалов для оцениваемых параметров

### Вариант 1

По данным налоговой службы, в регионе ведут деятельность 3546 индивидуальных предпринимателей. Для определения средней суммы ежегодных налоговых отчислений был произведен простой случайный отбор 60 налоговых деклараций индивидуальных предпринимателей. Получены следующие значения:

$$\sum_{i=1}^{1970} y_i = 3090449; \sum_{i=1}^{1970} y_i^2 = 165\,496\,084\,662.$$

Найдите доверительные границы, в которых с вероятностью 95% находится среднее значение ежегодных налоговых отчислений для индивидуальных предпринимателей региона.

### Вариант 2

Центр статистических исследований проводит анализ организаций некоторого региона, занятых в сфере образования. Аналитики измеряют количество сотрудников юридического отдела в таких организациях. Всего в регионе насчитывается 834 организации. В рамках исследования были взяты данные 70 компаний. Их данные приведены в таблице.

Сотр. Юр. Отдела	0	1	2	3	4	6
Кол-во орг.	3 6	1 9	1 0	2	2	1

Также известно, что в 484 организациях из 834 таких сотрудников нет вовсе.

Оцените общее количество юридических сотрудников, занятых в образовательной сфере, и постройте 95%-доверительный интервал для данного значения.

### Вариант 3

Некоторая ИТ-компания проводит онлайн- контрольную работу по математике для всех желающих. В работе 9 заданий. Всего КР написали 8651 человек. Для определения среднего числа решенных задач была произведена случайная выборка 70 решенных вариантов. Данные приведены в таблице.

Число решенных задач	0	1	2	3	4	5	6	7	8	9		
Количество людей	8	4	5	0	1	4	1	5	8	3	2	1

Также организаторы посмотрели на общее количество работ, выполненных, либо не выполненных, полностью:

Число решенных задач	0	9
Количество людей	101 1	1 32

## Тема Оценивание долей и процентов

### Вариант 1

Исследователи выясняют изменение популярности автомобилей с автоматической коробкой переключения передач в некотором городе. Известно, что 3 года назад доля новых проданных автомобилей с автоматической КПП составляла 42,76%. Для оценки изменений была произведена выборка 300 продаж текущего года. Данные приведены ниже:

*Оценка общего числа проданных новых автомобилей в городе за год: 53 000*

*Объем выборки: 300*

*Из них с АКПП: 143*

Основываясь на данных выборки, оцените вероятность того, что доля новых автомобилей с АКПП превышает значение, полученное 3 года назад.

### Вариант 2

В простой случайной выборке 200 колледжей из совокупности в 2000 колледжей 120 одобрили некоторое предложение, 57 были против и 23 колледжа не выразили своего мнения. Укажите 95%-ные доверительные границы для числа колледжей в совокупности, одобряющих это предложение.

### Вариант 3

Простая случайная выборка объемом в 290 домохозяйств извлечена из городского района, содержащего 14 828 домохозяйств. Каждой семье задавались вопросы: владеет ли она домом или нанимает квартиру и пользуется ли она отдельным туалетом. Результаты имеют вид:

		Домовладельцы	Нанимающий квартиру	Число семей
Пользование отдельным туалетом	Да	141	109	250
	Нет	6	34	40
				290

Для семей, нанимающих квартиру, оцените процент пользующихся отдельным туалетом и укажите стандартную ошибку вашей оценки.

**Тема** Критерии однородности вида смещение: классический критерий Стьюдента, критерий Вилкоксона

### Вариант 1

Центр статистики небольшого города собрал данные о зарплатах рабочих двух городских заводов. 12 рабочих завода А получают следующую зарплату: 24323, 33438, 21166, 29286, 29208, 30190, 28886, 30618, 38075, 29896, 39466, 36072 руб/мес, а 9 рабочих завода В – 24363, 31059, 32724, 29874, 33007, 30060, 33194, 31917, 26354 руб/мес. Верно ли предположение о равенстве средних зарплат рабочих заводов А и В? Примите уровень значимости 0,05.

### Вариант 2

В таблице приведены заработные платы случайно выбранных жителей некоторого города до и после проведения экономических реформ со стороны местной администрации. Изменение средней заработной платы – один из ключевых показателей их эффективности. На уровне значимости 0,05 оцените, является ли разница в средней заработной плате статистически значимой, или нет. Известно, что данные имеют нормальное распределение и дисперсии выборок одинаковы.

До реформ	После реформ
1695 7	23563
2211 2	38668
2593 2	35144
3222 0	34119
3007 4	34454
2816 8	40065
3076 3	24994
2794 7	34405
3276 1	42410
3154 9	35544
	28027
	41374

### Вариант 3

В рамках исследования влияния демографических показателей на заработную плату по определенным должностям, были исследованы вакансии компаний некоторого города на должность администратора в небольших магазинах торговых центров. Данные о предлагаемых зарплатах приведены ниже (в тысячах рублей в месяц). Верно ли предположение о том, что администраторам-девушкам в среднем предлагают несколько большую зарплату, чем администраторам мужчинам? Примите уровень значимости 0,05.

Вакансии, требующие девушек; т.р.	Вакансии, требующие мужчин; т.р.
47	58,5
53	48
60	50
61,5	54
48	60
54	49
56	46
62	52
60	50
	48
	56

**Тема** Комплексные задачи: Классический критерий Стьюдента, критерий однородности вида растяжение Фишера

### Вариант 1

В лаборатории провели сравнительный анализ двух приборов измерения количества веществ в газах. После 10 измерений прибором А были получены следующие результаты (в процентах общего объема газа): 90,86; 91,53; 91,69; 91,61; 91,45; 91,87; 91,53; 91,76; 91,8; 91,53. Измерения того же газа прибором Б проводились 11 раз: 91,19; 91,52; 91,22; 91,29; 91,45; 91,58; 91,34; 91,46; 91,69; 91,49. Проверьте гипотезу о том, что прибор Б является более точным, чем прибор А, учитывая, что значения имеют нормальное распределение. Примите уровень значимости 0,05.

### Вариант 2

Интернет-магазин бытовой техники исследует влияние рекламной кампании на продажу настольных часов. Для этого в течение 21 дня были измерены ежедневные продажи часов до рекламной кампании (СВ X, количество проданных часов за день), и в течение 28 дней замерялись ежедневные продажи после старта рекламной кампании (СВ Y, количество проданных часов за день). Агрегированные результаты приведены ниже. Проверьте на уровне значимости 0,05, что рекламная кампания положительно сказалась на продажах настольных часов, учитывая, что данные имеют нормальное распределение.

$$\sum_i x_i = 113; \sum_i x_i^2 = 759;$$
$$\sum_i y_i = 246; \sum_i y_i^2 = 2376;$$

### Вариант 3

Компания, владеющая ресторанами, работающими в двух городах, осуществляет бесплатную для клиентов доставку своей еды в офисы во время обедов. Для самой компании такие доставки являются платными – вознаграждение курьерам, проездные билеты, бензин и пр. Точная стоимость 1 доставки рассчитывается в конце рабочего дня – общая стоимость доставок, деленная на число доставок. Из дня в день она варьируется в силу разных причин – количество заказов на 1 курьера, пробки, удаленность клиентов и пр. В таблице ниже приведены данные о средней стоимости 1 доставки (в рублях) в случайные дни в двух городах. Верно ли предположение о том, что в городе Б спрогнозировать бюджет отдела доставки легче, чем в городе А, учитывая, что данные имеют нормальное распределение. Примите уровень значимости 0,05.

род А	Го	1	1	1	1	1	1	1	1	1	1
		26	61	49	59	38	85	63	66	21	42
род Б	Го	1	1	1	1	1	1	1	1	1	1
		51	81	67	68	65	68	59	67	53	66
род Б	Го	1	1	1	1	1					
		79	78	68	73	61					



## **Комплект заданий для выполнения лабораторных работ**

по дисциплине Анализ данных

**Лабораторная работа №1.** Применение пакетов статистического анализа для проверки гипотез об однородности вида смещение в двухвыборочной модели.

Задание: с использованием статистического пакета найти решение указанных задач; подготовить отчет.

Для каждой задачи:

- Сформулировать постановку задачи в математическом смысле
- Осуществить загрузку данных
- Выполнить предобработку данных
- Применить соответствующие программные компоненты для решения задачи
- Включить описание решение задачи в отчет о выполнении лабораторной работы (постановка задачи, описание ее решения, результаты и выводы)

Примерный перечень задач:

1. Станок штампует детали, размер которых соответствует заданному нормативу, т.е. вероятность превышения и занижения нормативного размера одинакова. Технологи провели наладку станка для того, чтобы уменьшить отклонения размеров изготовленных деталей от размера, требуемого стандартом. До и после наладки случайным образом было выбрано по 100 деталей. Можно ли считать, опираясь на представленные данные, что точность изготовления деталей увеличилась после наладки станка? Уровень значимости считать равным 0,05.
2. В таблице приведены заработные платы случайно выбранных жителей некоторого города до и после проведения экономических реформ со стороны местной администрации. Изменение средней заработной платы – один из ключевых показателей их эффективности. На уровне значимости 0,05 оцените, является ли разница в средней заработной плате статистически значимой, или нет. Известно, что данные имеют нормальное распределение, а дисперсии выборок можно считать одинаковыми.

**Лабораторная работа №2.** Применение пакетов статистического анализа для проверки гипотез об однородности в двухвыборочной модели (в общем случае).

Задание: с использованием статистического пакета найти решение указанных задач; подготовить отчет.

Для каждой задачи:

- Сформулировать постановку задачи в математическом смысле
- Осуществить загрузку данных
- Выполнить предобработку данных
- Применить дополнительные статистические методы для проверки промежуточных гипотез
- Применить соответствующие программные компоненты для решения задачи
- Включить описание решение задачи в отчет о выполнении лабораторной работы (постановка задачи, описание ее решения, результаты и выводы)

Примерный перечень задач:

1. Интернет-магазин бытовой техники исследует влияние рекламной кампании на продажу настольных часов. Для этого в течение 21 дня были измерены ежедневные продажи часов до рекламной кампании, и в течение 28 дней замерялись

ежедневные продажи после старта рекламной кампании. Данные приведены в таблице. Проверьте, что рекламная кампания положительно сказалась на продажах настольных часов.

2. Аналитики мессенджера сравнивают популярность функции общения в группах (чатах с большим количеством пользователей) в 2 странах СНГ. Для этого они случайным образом отобрали по 1500 групповых чатов из каждой страны, и замерили количество подключенных пользователей (принадлежность группы той или иной стране определялась преваляцией участников из рассматриваемой страны (более 70%)). Данные приведены в таблице. Оцените, различается ли среднее количество пользователей в групповых чатах между странами.

**Лабораторная работа №3.** Применение пакетов статистического анализа для исследования зависимости между количественными признаками.

Задание: с использованием статистического пакета найти решение указанных задач; подготовить отчет.

Для каждой задачи:

- Сформулировать постановку задачи в математическом смысле
- Осуществить загрузку данных
- Выполнить предобработку данных
- Применить дополнительные статистические методы для проверки промежуточных гипотез
- Применить соответствующие программные компоненты для решения задачи
- Включить описание решения задачи в отчет о выполнении лабораторной работы (постановка задачи, описание ее решения, результаты и выводы)

Примерный перечень задач:

1. Имеются данные о 64 подопытных мышах: содержание красных кровяных телец (RBC) и уровень гемоглобина (HGB). Проверить гипотезу о независимости параметров против альтернативы, что между ними существует положительная монотонная корреляция, с использованием критерия Спирмена.
2. Лабораторным прибором, определяющим содержание азота, был проведен многократный анализ 10 различных веществ. Из различных проб одного и того же вещества в итоге было вычислено среднее значение (среднее содержание азота в веществе, в %) и среднее отклонение (среднее отклонение измерений прибора, в %). Данные приведены в таблице. Проверить гипотезу о независимости точности измерений от процентного содержания азота, считая, что данные имеют нормальное распределение. Примите уровень значимости равным 0,01.
3. Имеются данные о 65 подопытных мышах: количество моноцитов (MO) и уровень протеина (TP). Проверить гипотезу о независимости параметров с использованием критерия Кендалла. Принять уровень значимости 0,01.

**Лабораторная работа №4.** Применение пакетов статистического анализа для исследования зависимости между количественными и номинальными признаками.

Задание: с использованием статистического пакета найти решение указанных задач; подготовить отчет.

Для каждой задачи:

- Сформулировать постановку задачи в математическом смысле
- Осуществить загрузку данных
- Выполнить предобработку данных
- Исследовать данные графическими методами (при необходимости)
- Применить дополнительные статистические методы для проверки промежуточных гипотез
- Применить соответствующие программные компоненты для решения задачи

- Рассчитать меры силы связей (при необходимости)
- Включить описание решение задачи в отчет о выполнении лабораторной работы (постановка задачи, описание ее решения, результаты и выводы)

Примерный перечень задач:

1. У каждого из 6800 респондентов измерялось два признака. Признак А (цвет глаз) имеет три градации: карий, серый, зеленый. Признак В (цвет волос) имеет градации: брюнет, шатен, блондин. Данные представлены в таблице. Имеется ли статистическая зависимость между признаками А и В на уровне значимости 0,05? Оцените силу связи в случае наличия зависимости.
2. В ходе одного из исследований были получены данные о весе 89 мужчин с одинаковым ростом 178 см. Вес указывается в зависимости от возраста опрашиваемого человека. Проверить гипотезу о независимости веса от возраста, применяя критерий Хи-квадрат. Предполагается, что молодые и пожилые респонденты имеют несколько меньший вес, чем респонденты в зрелом возрасте. Уровень значимости принять 0,01.
3. В одной из европейских стран была собрана статистика по росту населения. Данные статистики приведены в таблице в зависимости от пола респондента. Верно ли предположение о том, что есть зависимость между ростом респондента и его полом?

## Примерный перечень тестовых заданий

по дисциплине Анализ данных

- Квантилью случайной величины  $X$  уровня  $0,4$ , имеющей некоторое непрерывное распределение, называется такое число  $x$ , что
  - 40 % всех значений СВ  $X$  находятся левее значения  $x$  (т.е. они меньше  $x$ )
  - 60 % всех значений СВ  $X$  находятся правее значения  $x$  (т.е. они больше  $x$ )
  - 60 % всех значений СВ  $X$  находятся левее значения  $x$  (т.е. они меньше  $x$ )
  - 40 % всех значений СВ  $X$  находятся правее значения  $x$  (т.е. они больше  $x$ )
  - 0,4 % всех значений СВ  $X$  находятся левее значения  $x$  (т.е. они меньше  $x$ )
- Метод оценивания называется состоятельным, если
  - оценка становится в точности равной оцениваемому значению для совокупности при  $n = N$ .
  - среднее значение оценки, взятое по всем возможным выборкам данного объема  $n$ , в точности равно истинному значению для совокупности.
  - оценка в точности равна оцениваемому значению вне зависимости от элементов, попадающих в выборку.
- Из генеральной совокупности объемом 2500 элементов была отобрана выборка размером 40 единиц. Чему равна доля отбора?
- Вычислите ранг элемента выборки, имеющего значение 5 (выделен жирным, подчеркнутым):  
1, 1, 2, 3, 3, 4, 5, **5**, 6, 6, 8, 9, 10, 10, 12.
- Верно ли следующее утверждение:  
Чем больше размер выборки, тем точнее оценка и уже доверительный интервал. Ширина интервала и объем выборки зависят между собой линейно (то есть, увеличивая объем выборки в 2 раза, мы сокращаем доверительный интервал также в 2 раза).
- Для корректного применения классического критерия Стьюдента необходимо выполнение ряда предположений о данных. Укажите требуемые предположения:
  - Нормальность распределения выборок
  - Равенство средних значений
  - Равенство дисперсий
  - Независимость выборок
  - Одинаковое количество наблюдений в выборках
- Верно ли следующее утверждение:  
Чем больше величина ошибки I рода ( $\alpha$ ), тем меньше величина ошибки II рода ( $\beta$ ).
- Попадание значения статистики в критическую область, построенную для заданного уровня значимости  $\alpha$ , говорит о том, что
  - Нулевая гипотеза не отвергается
  - Нулевая гипотеза отвергается в пользу альтернативной гипотезы
- Имеется следующая задача:  
Согласно опросам 29 семей, проводившимся в 1968 г. в юго-западном регионе Англии, выборочное среднее еженедельной арендной платы за меблированную квартиру составило 2,5 £, а выборочная дисперсия 0,67 £<sup>2</sup>. В Уэльсе выборочное среднее арендной платы 16 семей составило 2,06 £, а выборочная дисперсия 0,42 £<sup>2</sup>. Проверьте, является ли различие арендной платы в этих регионах

статистически значимым, предполагая, что выборки порождены гауссовскими случайными величинами. Примите уровень значимости равным 0,05.

Какому критерию следует отдать предпочтение при решении данной задачи?

- Применить критерий Вилкоксона
- Сначала применить критерий Фишера для проверки равенства дисперсий, затем (если равны) – Стьюдента или Вилкоксона (если не равны)
- Применить критерий Стьюдента
- Применить критерий Спирмена
- Выборки  $X_m$  и  $Y_n$  называются однородными, если
  - $\exists t \in R^1: F_X(t) = F_Y(t)$
  - $F_X(t) = F_Y(t) \forall t \in R^1$
  - $\exists t \in R^1: F_X(t) \neq F_Y(t)$
  - $F_X(t) \neq F_Y(t) \forall t \in R^1$
- Для исследования зависимости между полом респондента и размером его заработной платы на некоторой позиции следует отдать предпочтение
  - Критерию Кендалла
  - Критерию Спирмена
  - Критерию, основанному на выборочном коэффициенте корреляции
  - Критерию хи-квадрат
- Укажите правильные границы возможных значений коэффициента согласованности Кендалла
  - $[0; 1]$
  - $(-1; 1)$
  - $[-1; 1]$
  - $(0; 1]$
- Верно ли следующее утверждение:  
Критерии Кендалла и Спирмена подходят только для определения монотонных зависимостей, и могут оказаться неспособны уловить немонотонные зависимости.
- В чем состоит основное отличие между мерами силы связи Пирсона и Крамера
  - Мера Крамера всегда больше меры Пирсона
  - Мера Крамера не зависит от объема выборки
  - Значение меры Крамера может достигать единицы
- Какие из следующих критериев являются ранговыми (работают с рангами наблюдений, а не с самими наблюдениями)
  - Критерий, основанный на выборочном коэффициенте корреляции
  - Критерий Кендалла
  - Критерий Спирмена