

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ
Заведующий кафедрой
Математических методов исследования операций
Азарнова Т.В.
22.03.2024



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Б1.О.07. Статистический анализ данных

1. Код и наименование направления подготовки/специальности:

09.04.03 Прикладная информатика

2. Профиль подготовки/специализация: Прикладная информатика в социальных и медицинских системах

3. Квалификация выпускника: магистр

4. Форма обучения: очная

5. Кафедра, отвечающая за реализацию дисциплины: Математических методов исследования операций

6. Составители программы: Азарнова Татьяна Васильевна, доктор техн. наук, профессор

7. Рекомендована: НМС факультета Прикладной математики информатики и механики, протокол № 5 от 22.03.2024

8. Учебный год: 2024/2025

Семестр(-ы): 1

9. Цели и задачи учебной дисциплины:

Цель: В рамках данного курса слушатели получают знания о математическом аппарате анализа статистических данных различной природы и приобретают навыки в математическом моделировании процесса исследования, т.е. в искусстве формализации постановки реальной задачи, которое заключается в умении перевести задачу с языка проблемно-содержательного (медицинского, технического и т.п.) на язык абстрактных математических схем и моделей

Задачи дисциплины – формирование знаний, умений и навыков по следующим направлениям: способы организации выборок; методы проверки статистических гипотез; регрессионный анализ; дисперсионный анализ; факторный анализ; методы классификации; дискриминантный анализ; деревья решений; анализ временных рядов.

10. Место учебной дисциплины в структуре ООП:

Дисциплина относится к обязательным дисциплинам базового цикла.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

| Код | Название компетенции | Код(ы) | Индикатор(ы) | Планируемые результаты обучения |
|-------|---|----------|---|---|
| ОПК-1 | Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте; | ОПК-1.2; | ОПК-1.2 Осуществляет первичный сбор и анализ материала, чтобы интерпретировать различные математические и информационные объекты | знать: – методы формирования выборок; – методы предварительной обработки данных; – методы отбора информативных признаков; – методы проверки статистических гипотез; – методы оценки зависимости между категоризованными, порядковыми и интервальными данными; – методы дисперсионного анализа; – методы факторного анализа; – методы классификации; – методы регрессионного анализа; – методы анализа временных рядов; уметь: – проводить опросы и анкетирование для сбора информации; – анализировать многомерные данные и формировать систему методов их обработки для интерпретации различных математических и информационных объектов; владеть (иметь навык(и)): – подбора методов сбора |

| | | | | |
|-------|--|----------------------|--|--|
| | | | | <p>информации и проверки качества моделей обработки данных в задачах исследования информационных процессов;</p> <ul style="list-style-type: none"> – интерпретации полученных результатов в терминах прикладной области с целью получения новых знаний и выводов. |
| ОПК-3 | <p>Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациям</p> | ОПК-3.1; ОПК-3.2; | <p>ОПК-3.1 Демонстрирует владение принципами сбора и анализа профессиональной информации, в том числе, с помощью статистических методов анализа.</p> <p>ОПК-3.2 Применяет новые математические методы анализа, визуализации и обработки различных типов информации для подготовки аналитических обзоров с выводами и рекомендациями.</p> | <p>знать:</p> <ul style="list-style-type: none"> – методы организации сбора информации, необходимой для решения исследовательских задач, поставленных специалистом более высокой квалификации; – статистические методы анализа информации для решения поставленных научно-прикладных задач; <p>уметь:</p> <ul style="list-style-type: none"> – применять методы статистического анализа данных для решения поставленных задач; – использовать возможности пакета прикладных программ для обработки информации и решения поставленных задач; <p>владеть (иметь навык(и)):</p> <ul style="list-style-type: none"> - методами анализа, визуализации и обработки различных типов информации для подготовки аналитических обзоров с выводами и рекомендациями. |

| | | | | |
|------|------------------------------------|--------|---|---|
| ПК-3 | Управление аналитическими работами | ПК-3.2 | Применяет математические методы для совершенствования различных инструментов исследований и выполнения аналитических работ. | <p>знать:</p> <ul style="list-style-type: none"> – методы проверки статистических гипотез; – методы оценки зависимости между категоризованными, порядковыми и интервальными данными; – методы дисперсионного анализа; – методы факторного анализа; – методы классификации; – методы регрессионного анализа; – методы анализа временных рядов; <p>уметь:</p> <ul style="list-style-type: none"> – применять математические методы для совершенствования различных инструментов исследований и выполнения аналитических работ.; <p>владеть (иметь навык(и)):</p> <ul style="list-style-type: none"> – методами проектирования и разработки информационных технологий для совершенствования различных инструментов исследований и выполнения аналитических работ. |
|------|------------------------------------|--------|---|---|

12. Объем дисциплины в зачетных единицах/час (в соответствии с учебным планом) — 4/144.

Форма промежуточной аттестации(зачет/экзамен) экзамен

13. Трудоемкость по видам учебной работы

| Вид учебной работы | | Трудоемкость | |
|--|--------------|--------------|--------------|
| | | Всего | По семестрам |
| | | | № семестра 1 |
| Контактная работа | | | |
| в том числе: | лекции | 32 | 32 |
| | практические | 16 | 16 |
| | лабораторные | 16 | 16 |
| | | | |
| Самостоятельная работа в том числе: курсовая работа (проект) | | 76 | 76 |
| Форма промежуточной аттестации | | 36 | 36 |
| Итого: | | 144 | 144 |

13.1. Содержание дисциплины

| п/п | Наименование раздела дисциплины | Содержание раздела дисциплины | Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК * |
|------------------|---|--|--|
| 1. Лекции | | | |
| 1 | Первичная статистическая обработка данных | Шкалы измерений. Унифицированное представление разнотипных данных. Способы заполнения пропущенных данных. Визуализация многомерных данных. Анализ резко выделяющихся наблюдений | Анализ данных |
| 2 | Проверка статистических гипотез в прикладных задачах | Основные типы гипотез, проверяемых в ходе статистического анализа и моделирования. Критерий знаков для одной выборки. Критерий Манна-Уитни. Критерий Уилкоксона. Критерий знаков для анализа парных повторных наблюдений. Критерий знаковых ранговых сумм Уилкоксона. Проверка гипотез, связанных с параметрами нормального распределения. | Анализ данных |
| 3 | Дисперсионный анализ | Однофакторный дисперсионный анализ Многофакторный дисперсионный анализ Ранговый дисперсионный анализ | Анализ данных |
| 4 | Анализ структуры и тесноты статистической связи между исследуемыми переменными. | Анализ тесноты связи между количественными переменными. Анализ статистической связи между порядковыми переменными. Анализ связей между классификационными переменными. | Анализ данных |
| 5 | Факторный анализ | Факторный анализ при жестко фиксированном количестве факторов Факторный анализ при нефиксированном количестве факторов Методы вращения и интерпретации факторов | Анализ данных |
| 6 | Распознавание образов и типологизация объектов в социально – экономических исследованиях. | Сущность, типологизация и прикладная направленность задач классификации объектов. Классификация при наличии обучающих выборок (дискриминантный анализ). Классификация без обучения (параметрический случай): расщепление смесей вероятностных распределений. Классификация без обучения | Анализ данных |

| | | | |
|--------------------------------|---|---|---------------|
| | | (непараметрический случай): кластер-анализ. | |
| 7 | Временные ряды | <p>Понятие временного ряда. Компоненты временных рядов. Стационарные временные ряды и их основные характеристики. Неслучайная составляющая временного ряда и методы его сглаживания. Модели нестационарных временных рядов и их идентификация. Прогнозирование экономических показателей, основанное на использовании временных рядов.</p> | Анализ данных |
| 2. Лабораторные занятия | | | |
| 1 | Первичная статистическая обработка данных | <p>Анализ данных в пакете статистического анализа данных по темам: -Шкалы измерений. -Унифицированное представление разнотипных данных. -Способы заполнения пропущенных данных. -Визуализация многомерных данных. -Анализ резко выделяющихся наблюдений</p> | Анализ данных |
| 2 | Проверка статистических гипотез в прикладных задачах | <p>Анализ данных в пакете статистического анализа данных по темам: - Критерий знаков для одной выборки. -Критерий Манна-Уитни. -Критерий Уилкоксона. -Критерий знаков для анализа парных повторных наблюдений. -Критерий знаковых ранговых сумм Уилкоксона. -Проверка гипотез, связанных с параметрами нормального распределения.</p> | Анализ данных |
| 3 | Дисперсионный анализ | <p>Анализ данных в пакете статистического анализа данных по темам: -Однофакторный дисперсионный анализ. -Многофакторный дисперсионный анализ. -Ранговый дисперсионный анализ.</p> | Анализ данных |
| 4 | Анализ структуры и тесноты статистической связи между исследуемыми переменными. | <p>Анализ тесноты связи между количественными переменными в пакете статистического анализа данных. Анализ статистической связи между порядковыми переменными в пакете статистического анализа данных. Анализ связей между классификационными переменными в пакете статистического анализа данных.</p> | Анализ данных |
| 5 | Факторный анализ | Факторный анализ при жестко фиксированном количестве факторов в | Анализ данных |

| | | | |
|---|---|---|---------------|
| | | пакете статистического анализа данных. Факторный анализ при нефиксированном количестве факторов в пакете статистического анализа данных. Методы вращения и интерпретации факторов в пакете статистического анализа данных. | |
| 6 | Распознавание образов и типологизация объектов в социально – экономических исследованиях. | Классификация при наличии обучающих выборок (дискриминантный анализ) в пакете статистического анализа данных. Классификация без обучения (параметрический случай): расщепление смесей вероятностных распределений в пакете статистического анализа данных. Классификация без обучения (непараметрический случай): кластер-анализ в пакете статистического анализа данных. | Анализ данных |
| 7 | Временные ряды | Прогнозирование экономических показателей, основанное на использовании временных рядов в пакете статистического анализа данных. | Анализ данных |

13.2. Темы (разделы) дисциплины и виды занятий

| № п/п | Наименование темы (раздела) дисциплины | Виды занятий (часов) | | | | |
|-------|---|----------------------|--------------|--------------|------------------------|-------|
| | | Лекции | Практическое | Лабораторные | Самостоятельная работа | Всего |
| 1 | Первичная статистическая обработка данных | 2 | | 2 | 10 | 14 |
| 2 | Проверка статистических гипотез в прикладных задачах | 2 | | 2 | 10 | 14 |
| 3 | Дисперсионный анализ | 2 | | 2 | 10 | 14 |
| 4 | Анализ структуры и тесноты статистической связи между исследуемыми переменными. | 4 | | 2 | 10 | 16 |
| 5 | Факторный анализ | 2 | | 2 | 10 | 14 |
| 6 | Распознавание образов и типологизация объектов в социально – экономических исследованиях. | 2 | | 2 | 10 | 14 |
| 7 | Временные ряды | 2 | | 4 | 16 | 22 |
| | Итого | 16 | | 16 | 76 | 108 |

14. Методические указания для обучающихся по освоению дисциплины

Для лучшего усвоения материала студентам рекомендуется домашняя работа с конспектами лекций, презентациями, выполнение практических заданий для самостоятельной работы, выполнение лабораторных работ, использование рекомендованной литературы и методических материалов. В рамках общего объема часов, отведенных для изучения дисциплины, предусматривается выполнение следующих видов самостоятельных работ студентов (СРС): изучение теоретического материала, выполнение в пакете статистического анализа данных заданий по темам, изученным на лекционных и практических занятиях.

При использовании дистанционных образовательных технологий и электронного обучения выполнять все указания преподавателей, вовремя подключаться к online занятиям, ответственно подходить к заданиям для самостоятельной работы.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины (список литературы оформляется в соответствии с требованиями ГОСТ и используется общая сквозная нумерация для всех видов источников)

а) основная литература:

| № п/п | Источник |
|-------|---|
| 1 | Котиков, П. Е. Анализ данных : учебно-методическое пособие / П. Е. Котиков. — Санкт-Петербург : СПбГПМУ, 2019. — 48 с. — ISBN 978-5-907184-46-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/174498 |
| 2 | Вольфсон, М. Б. Анализ данных : учебное пособие / М. Б. Вольфсон. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2015. — 81 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/180254 |
| 3 | Кузьмин, В. И. Методы анализа данных : учебное пособие / В. И. Кузьмин, А. Ф. Гадзаов. — 2-е изд., перераб. и доп. — Москва : РТУ МИРЭА, 2020. — 155 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/171433 . |

б) дополнительная литература:

| № п/п | Источник |
|-------|--|
| 4 | Горелов, В. И. Анализ статистических данных : практикум : [16+] / В. И. Горелов, Т. Н. Ледащева ; Российская международная академия туризма. — Москва : Университетская книга, 2015. — 120 с. : ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=574944 . — Библиогр.: с. 107. — ISBN 978-5-98699-151-1. — Текст : электронный. |
| 5 | Агалаков, С. А. Статистические методы анализа данных : учебное пособие : [16+] / С. А. Агалаков ; Омский государственный университет им. Ф. М. Достоевского. — Омск : Омский государственный университет им. Ф.М. Достоевского, 2017. — 92 с. : табл., граф., схем., ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=562918 . — Библиогр. в кн. — ISBN 978-5-7779-2187-1. — Текст : электронный. |
| 6 | Жуковский, О. И. Информационные технологии и анализ данных : учебное пособие / О. И. Жуковский ; Томский Государственный университет систем управления и радиоэлектроники (ТУСУР). — Томск : Эль Контент, 2014. — 130 с. : схем., ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=480500 . — Библиогр.: с. 126. — ISBN 978-5-4332-0158-3. — Текст : электронный. |

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

| № п/п | Ресурс |
|-------|-------------------|
| 1. | <i>ЭБС Лань</i> |
| 2. | <i>ЭБС ЮРАЙТ</i> |
| 3. | <i>edu.vsu.ru</i> |

16. Перечень учебно-методического обеспечения для самостоятельной работы (учебно-методические рекомендации, пособия, задачки, методические указания по выполнению практических (контрольных), курсовых работ и др.)

| № п/п | Источник |
|-------|--|
| 1 | Зубов, Н. Н. Статистика в биомедицине, фармации и фармацевтике : учебное пособие : [16+] / Н. Н. Зубов, В. И. Кувакин, С. З. Умаров ; под общ. ред. И. А. Наркевича. – Москва ; Берлин : Директ-Медиа, 2019. – 386 с. : ил., табл. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=578236 . – Библиогр.: с. 326-327. – ISBN 978-5-4499-1173-5. – DOI 10.23681/578236. – Текст : электронный. |
| 2 | 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. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=227280 . – ISBN 978-5-7598-0866-4. – Текст : электронный. |

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение):

При реализации дисциплины могут использоваться технологии электронного обучения и дистанционные образовательные технологии на базе портала edu.vsu.ru, а также другие доступные ресурсы сети Интернет.

18. Материально-техническое обеспечение дисциплины:

Лекции: лекционная аудитория, учебная мебель, компьютер (ноутбук), мультимедийное оборудование (проектор, экран, средства звуковоспроизведения).

Практические и лабораторные занятия: специализированная аудитория, оснащенная учебной мебелью и персональными компьютерами для индивидуальной работы с возможностью подключения к сети «Интернет» (компьютерные классы, студии), мультимедийное оборудование (проектор, экран, средства звуковоспроизведения).

Самостоятельная работа: учебная мебель, компьютерный класс, компьютер с возможностью подключения к сети «Интернет» и платформе Электронного университета ВГУ (LMS moodle).

Программное обеспечение:

- ОС Windows 8 (10),
- интернет-браузер (Mozilla Firefox);
- ПО Adobe Reader;
- пакет стандартных офисных приложений для работы с документами, таблицами (МойОфис, LibreOffice);
- специализированное ПО, допускается демоверсия или виртуальный аналог ПО.

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

| № п/п | Наименование раздела дисциплины (модуля) | Компетенция(и) | Индикатор(ы) достижения компетенции | Оценочные средства |
|---|--|------------------------|---|-----------------------------------|
| 1. | Первичная статистическая обработка данных | ОПК-1 ОПК-3 ПК-3 | ОПК-1.2 ОПК-3.1 ОПК-3.2 ПК-3.2 | Задание для лабораторной работы 1 |
| 2. | Проверка статистических гипотез в прикладных задачах | ОПК-1 ОПК-3 ПК-3 | ОПК-1.2 ОПК-3.1 ОПК-3.2 ПК-3.2 | Задание для лабораторной работы 1 |
| 3 | Дисперсионный анализ | ОПК-1 ОПК-3 ПК-3 | ОПК-1.2 ОПК-3.1 ОПК-3.2 ПК-3.2 | Задание для лабораторной работы 1 |
| 4 | Анализ структуры и тесноты статистической связи между исследуемыми переменными. | ОПК-1 ОПК-3 ПК-3 | ОПК-1.2 ОПК-3.1 ОПК-3.2 ПК-3.2 | Задание для лабораторной работы 1 |
| 5 | Факторный анализ | ОПК-1 ОПК-3 ПК-3 | ОПК-1.2 ОПК-3.1 ОПК-3.2 ПК-3.2 | Задание для лабораторной работы 2 |
| 6 | Распознавание образов и типологизация объектов социально-экономических исследований. | ОПК-1 ОПК-3 ПК-3 | ОПК-1.2 ОПК-3.1 ОПК-3.2 ПК-3.2 | Задание для лабораторной работы 3 |
| 7 | Временные ряды | ОПК-1 ОПК-3 ПК-3 | ОПК-1.2 ОПК-3.1 ОПК-3.2 ПК-3.2 | Задание для лабораторной работы 4 |
| Промежуточная аттестация форма контроля - экзамен | | | | Тест |

20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

Тестовые задания, Лабораторные работы, Устный опрос

1. Лабораторная работа №1

1. В предложенном вам файле «Таблица квартир» рассмотрите переменную Цена. Вычислите все описательные статистики и постройте простейшие статистические графики для данной переменной, проверьте гипотезу о нормальности, используя глазомерный метод проверки на нормальность. Проинтерпретируйте полученные результаты.
2. В предложенном вам файле «Cars» рассмотрите переменную Цена. Вычислите все описательные статистики и постройте простейшие статистические графики для данной переменной, проверьте гипотезу о нормальности, используя глазомерный метод проверки на нормальность. Проинтерпретируйте полученные результаты.
3. В предложенном вам файле «Tights» рассмотрите переменную Цена. Вычислите все описательные статистики и постройте простейшие статистические графики для данной переменной, проверьте гипотезу о нормальности, используя глазомерный метод проверки на нормальность. Проинтерпретируйте полученные результаты.
4. В предложенном вам файле «Таблица квартир» рассмотрите переменную Цена. Вычислите все описательные статистики и постройте простейшие статистические графики для данной переменной, проверьте гипотезу о нормальности, используя глазомерный метод проверки на нормальность. Проинтерпретируйте полученные результаты.
5. В предложенном вам файле «Продолжительность жизни» рассмотрите переменную «в городе». Вычислите все описательные статистики и постройте простейшие статистические графики для данной переменной, проверьте гипотезу о нормальности, используя глазомерный метод проверки на нормальность. Проинтерпретируйте полученные результаты.
6. В предложенном вам файле «Продолжительность жизни» рассмотрите переменную «в селе». Вычислите все описательные статистики и постройте простейшие статистические графики для данной переменной, проверьте гипотезу о нормальности, используя глазомерный метод проверки на нормальность. Проинтерпретируйте полученные результаты.
7. В предложенном вам файле «Продолжительность жизни» рассмотрите переменную «в селе». Вычислите все описательные статистики и постройте простейшие статистические графики для данной переменной, проверьте гипотезу о нормальности, используя глазомерный метод проверки на нормальность. Проинтерпретируйте полученные результаты.
8. В предложенном вам файле Adstudy (стандартные примеры) рассмотрите переменные «Gender» и «Advert». Проверьте гипотезу о независимости этих двух признаков.
9. В предложенном вам файле «школьники» рассмотрите переменные «Continuation decision» и «I for girls, 0 for boys». Проверьте гипотезу о независимости этих двух признаков.
10. В предложенном вам файле «школьники» рассмотрите переменные «Continuation decision» и «Number of younger siblings (at age 16)». Проверьте гипотезу о независимости этих двух признаков.
11. В предложенном вам файле «Качество работы» рассмотрите переменные «Приоритеты в работе» и «Состав семьи». Проверьте гипотезу о независимости этих двух признаков.
12. В предложенном вам файле «1960-1985» рассмотрите переменные «NOIL» и «INTER». Проверьте гипотезу о независимости этих двух признаков.
13. В предложенном вам файле NEW11 приведены данные об объемах продаж в 15 магазинах до и после рекламы. Выяснить, значим ли эффект рекламы.
14. С помощью t-теста для независимых признаков проверьте гипотезу о том, средние значения переменной «Number of O-levels obtained at national exams at age 16 (prior to continuation decision)» одинаковы для мальчиков и девочек. Файл «школьники».
15. С помощью t-теста для независимых признаков проверьте гипотезу о том, средние значения переменной «price» одинаковы для различных значений переменной «foreign/russian». Файл «Mydata».
16. С помощью t-теста для независимых признаков проверьте гипотезу о том, средние значения переменной «price» одинаковы для различных значений переменной «status». Файл «Mydata».

17. С помощью t-теста для независимых признаков проверьте гипотезу о том, средние значения переменной « Number of O-levels obtained at national exams at age 16 (prior to continuation decision)» одинаковы для мальчиков и девочек. Файл «школьники».
18. С помощью t-теста для независимых признаков проверьте гипотезу о том, средние значения переменной «цена » одинаковы для различных фирм. Файл «Tight».
19. С помощью t-теста для зависимых признаков проверьте гипотезу о том, средние значения переменной «GDP60» и переменной «GDP85» равны (нет эффекта времени) Файл «1960-1985».
20. С помощью дисперсионного анализа проверьте влияние «округа» на «уровень безработицы». Файл «Экономика». Проинтерпретируйте полученные результаты.
21. С помощью дисперсионного анализа проверьте влияние «округа» на «коэффициент расслоения». Файл «Экономика». Проинтерпретируйте полученные результаты.
22. В предложенном вам файле Employees закодируйте переменную Age тремя значениями (молодые, средний возраст, зрелый возраст). Проверьте с помощью дисперсионного анализа влияние возраста на заработок (“SALARY”). Проинтерпретируйте полученные результаты.
23. С помощью дисперсионного анализа проверьте влияние «диагональ» на «цену». Файл «LCD». Проинтерпретируйте полученные результаты.
24. С помощью дисперсионного анализа проверьте влияние «TCO» на «цену». Файл «LCD». Проинтерпретируйте полученные результаты.
25. В предложенном вам файле «Mydata» закодируйте переменную «gun» тремя значениями (маленький, средний, большой). Проверьте с помощью дисперсионного анализа влияние пробега на цену. Проинтерпретируйте полученные результаты.

| | |
|---------------|---|
| NUM | номер страны в базе данных Summers, Heston (1988); |
| NOIL (dummy) | 1 для страны, не добывающей нефть, 0 - для добывающей; |
| INTER (dummy) | 1 для страны с хорошим качеством данных, 0 - в противном случае; |
| OECD (dummy) | 1 для страны |
| GDP60 | ВВП на душу населения в 1960 г.; |
| GDP85 | ВВП на душу населения в 1985 г.; |
| GDPGRO | средний рост ВВП на душу населения с 1960 г. по 1985 г. (в %); |
| POPGRO | средний рост работоспособного населения с 1960 г. по 1985 г. (в %); |
| IONY | средняя доля инвестиций (включая государственные) в общем объеме ВВП с 1960 г. по 1985 г. (в %); |
| SCH | средняя доля работоспособного населения, имеющая полное среднее образование с 1960 г. по 1985 г. (в %); |
| LIT | доля людей среди населения старше 15 лет, умеющих читать и писать в 1960 г. |

Лабораторная работа №2

1. В предложенном файле содержатся результаты четырех тестов для пожилых людей: arith – арифметический тест, info – информационный тест, picture - тест дополнения картинок, similars – тест на подобие. Провести факторный анализ на основе данной выборки. Оставить оптимальное количество факторов, оценить качество восстановления корреляционной матрицы, используя графический анализ и методы вращения дать интерпретацию полученных факторов через исходные переменные.

Лабораторная работа №3

Выполнить следующее задание по дискриминантному анализу в пакете прикладных программ Statistica :

1. Проверить предпосылки дискриминантного анализа для предложенных данных:
 - 1.1 . Нормальность распределения признаков внутри групп.
 - 1.2 . Совпадение ковариационных матриц во всех группах.
2. Построить матрицы межгруппового и внутригруппового рассеивания.
3. Построить дискриминантные функции.
4. По стандартизованным коэффициентам проинтерпретировать существенность признаков с точки зрения различия между классами.
5. По структурным коэффициентам проинтерпретировать существенность признаков с точки зрения различия между классами.
6. Оценить требуемое количество дискриминантных функций по критерию χ^2 .
7. Оценить качество дискриминантных функций по расположению классов в пространстве дискриминантных функций.
8. Построить классифицирующие функции Фишера.
9. Оценить качество классификации по классифицирующим функциям Фишера.
10. Оценить апостериорные вероятности принадлежности объектов к классам, априорные вероятности задать самостоятельно.
11. Определить принадлежность нового объекта к одному из трех классов.

Лабораторная работа №4

1. В таблице 1, приведенной ниже приведены данные по доходам населения за период с 1999 по 2001 год и сведения о приросте сбережений на вкладах и в ценных бумагах и о расходах на покупку валюты за тот же период.

Таблица 1.

| | Доходы населения, млрд. р. | Прирост сбережений во вкладах и ценных бумагах, млрд. р. | Расходы на покупку валюты, млрд. р. |
|---------------|----------------------------|--|-------------------------------------|
| Январь 1999 | 166,2 | 4,3 | 13,8 |
| Февраль 1999 | 186 | 8,2 | 13,6 |
| Март 1999 | 197,9 | 4,4 | 21,8 |
| Апрель 1999 | 220,5 | 9,3 | 15 |
| Май 1999 | 212,5 | 8,3 | 13,6 |
| Июнь 1999 | 226,5 | 10,2 | 17,4 |
| Июль 1999 | 226,6 | 8,8 | 21,3 |
| Август 1999 | 239,1 | 5,5 | 22,2 |
| Сентябрь 1999 | 239,8 | 6,2 | 20,4 |
| Октябрь 1999 | 250,8 | 7 | 18,1 |
| Ноябрь 1999 | 257 | 8,2 | 21,8 |
| Декабрь 1999 | 354,9 | 17,4 | 27,7 |
| Январь 2000 | 215 | 8,6 | 17,2 |
| Февраль 2000 | 261,3 | 12,3 | 17,5 |
| Март 2000 | 286,5 | 13,2 | 22,6 |

| | | | |
|---------------|-------|------|------|
| Апрель 2000 | 291,5 | 11,1 | 18,4 |
| Май 2000 | 284,5 | 15,6 | 16,5 |
| Июнь 2000 | 315,1 | 17 | 18,3 |
| Июль 2000 | 308,1 | 13,2 | 20,6 |
| Август 2000 | 322,7 | 9,4 | 23,9 |
| Сентябрь 2000 | 331,5 | 10,9 | 22,9 |
| Октябрь 2000 | 325,5 | 7,8 | 24,7 |
| Ноябрь 2000 | 348,5 | 15 | 22,3 |
| Декабрь 2000 | 452,3 | 8,1 | 28,9 |
| Январь 2001 | 290,2 | 13,3 | 20,5 |
| Февраль 2001 | 337,5 | 12,8 | 20,2 |
| Март 2001 | 376,1 | 15 | 21,3 |
| Апрель 2001 | 395,4 | 17 | 21,2 |
| Май 2001 | 372,1 | 11,2 | 22,6 |
| Июнь 2001 | 428,2 | 17,1 | 23,7 |
| Июль 2001 | 424,9 | 14,9 | 26,7 |
| Август 2001 | 437,2 | 16,2 | 29 |
| Сентябрь 2001 | 436,1 | 20,5 | 22,6 |
| Октябрь 2001 | 438,6 | 17,5 | 26,2 |
| Ноябрь 2001 | 448,3 | 17,9 | 31,3 |
| Декабрь 2001 | 580,6 | 22,4 | 35,4 |

Методом экспоненциального сглаживания постройте модель одного из предложенных временных рядов. Оцените качество модели и сделайте прогноз на несколько шагов вперед.

- Проведите сезонную декомпозицию для ряда Series_G, расположенного в папке Examples. Укажите: имеет ли предложенный временной ряд тренд, сезонную компоненту, вид сезонности, период сезонности. Проанализируйте, является ли случайная компонента данного ряда стационарным рядом.

Критерии оценки лабораторных работ:

- оценка «отлично» выставляется студенту, если все задания выполнены;
- оценка «хорошо» выставляется студенту, если все задания выполнены, но возможно, с некоторыми недочетами
- оценка «удовлетворительно» выставляется студенту, если задания выполнены частично и (или) с недочетами.
- оценка «неудовлетворительно», если выполнено меньше 50 % задания.

20.2 Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: тест.

Вопросы к экзамену:

1. Описательные статистики
2. Методы проверки распределения выборки на нормальность
3. Методы проверки статистических гипотез
4. Анализ таблиц сопряженности
5. Анализ статистической связи между порядковыми переменными.
6. t-тест для независимых выборок
7. t-тест для зависимых выборок
8. Однофакторный дисперсионный анализ
9. Ранговый дисперсионный анализ
10. Факторный анализ
11. Иерархические методы кластеризации
12. Метод k-средних
13. Дискриминантный анализ
14. Понятие временного ряда.
15. Компоненты временных рядов.
16. Стационарные временные ряды и их основные характеристики.
17. Проверка гипотез о наличии тренда
18. Метод сглаживания скользящими средними
19. Метод сезонной декомпозиции
20. Адаптивные методы прогнозирования
21. Метод экспоненциального сглаживания.

Тест.

Тест используется для оценки знаний студентов на экзамене и представляет контрольно-измерительный материал промежуточной аттестации, позволяющий оценить степень сформированности знаний, умений и навыков.

Для оценивания результатов обучения на экзамене используются следующие показатели:

- 1) знание учебного материала и владение понятийным аппаратом статистического анализа данных;
- 2) умение анализировать многомерные данные;
- 3) умение применять методы анализа данных при решении задач в различных прикладных областях;
- 4) владение навыками построения и проверки качества моделей анализа данных;
- 7) владение навыками интерпретации полученных результатов в терминах прикладной области с целью получения новых знаний и выводов.

Контрольно-измерительный материал №_1

1. Проинтерпретируйте основные статистические показатели, используемые в описательной статистике: *Minimum, maximum, Range, Медиана, Доверительный интервал, Variance, Lower, Upper Quartile, Std.Dev, Skewness, Kurtosis*. (1 балл).
2. Опишите t-тест для зависимых и независимых выборок. Приведите примеры задач на использование данных методов. (2 балл).
3. Приведите пример задачи на использование рангового дисперсионного анализа. (1 балл)
4. Опишите метод решения задачи дискриминации в дискриминантном анализе. (1 балл)

5. Опишите метод интерпретации факторов в факторном анализе (1 балл).
6. Сформулируйте алгоритм метода древовидной классификации. (2 балла).
7. Дайте определение временного ряда. Нарисуйте пример графика временного ряда. (1 балл)
8. Охарактеризуйте основные компоненты временных рядов. (1 балл).
9. Дайте определение автокорреляционной функции для стационарных временных рядов. (1 балл).
10. Опишите метод восходящих и нисходящих серий для проверки гипотезы о наличии тренда у ряда. (1 балл)
11. Сформулируйте алгоритм сглаживания скользящими средними для временных рядов. (2 балла).
12. Опишите подробно метод экспоненциального сглаживания для простейших временных рядов. (2 балла).

Контрольно-измерительный материал №_2

1. Проинтерпретируйте основные статистические показатели, используемые в описательной статистике: *Minimum, maximum, Range, Медиана, Доверительный интервал, Variance, Lower, Upper Quartile, Std.Dev, Skewness, Kurtosis*. (1 балл).
2. Методы проверки гипотезы о нормальном распределении признака. (1 балл).
3. Опишите общий принцип проверки статистических гипотез. (1 балл)
4. Опишите принцип построения таблиц сопряженности для проверки гипотезы о независимости двух категоризованных переменных. Приведите пример задачи на использование данного метода. (2 балл).
3. Как определяется количество дискриминантных функций в дискриминантном анализе. (1 балл)
4. Опишите последовательность шагов факторного анализа (1 балл).
6. Сформулируйте алгоритм метода к-средних. (2 балла).
7. Дайте определение временного ряда. Нарисуйте пример графика временного ряда. (1 балл)
8. Охарактеризуйте основные компоненты временных рядов. (1 балл).
9. Дайте определение автокорреляционной функции для стационарных временных рядов. (1 балл).
10. Опишите метод сезонной декомпозиции временных рядов. (2 балла)
11. Сформулируйте алгоритм сглаживания скользящими средними для временных рядов. (2 балла).
12. Опишите подробно метод экспоненциального сглаживания для простейших временных рядов. (2 балла).

Для оценивания результатов обучения на экзамене используется 4-балльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Соотношение показателей, критериев и шкалы оценивания результатов обучения.

| Критерии оценивания | Шкала оценок |
|---|--------------|
| Обучающийся в полной мере владеет понятийным аппаратом данной области науки (теоретическими основами дисциплины), сдал все практические и лабораторные работы, среднее количество правильных ответов на вопросы тестов превышает 80%. | Отлично |

| | |
|--|---------------------|
| Обучающийся владеет понятийным аппаратом данной области науки (теоретическими основами дисциплины), но не сдал одну практическую или лабораторную работу, среднее количество правильных ответов на вопросы тестов находится в диапазоне 70-80%. | Хорошо |
| Обучающийся демонстрирует неуверенное владение понятийным аппаратом данной области науки (теоретическими основами дисциплины), не сдал две практических или лабораторных работы, среднее количество правильных ответов на вопросы тестов находится в диапазоне 60-70%. | Удовлетворительно |
| Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки, не сдал более двух практических или лабораторных работы, среднее количество правильных ответов на вопросы тестов менее 70%. | Неудовлетворительно |

20.3 Фонд оценочных средств сформированности компетенций студентов, рекомендуемый для проведения диагностических работ

ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте

ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

Вопросы с выбором ответа

- Основная гипотеза t-теста для зависимых выборок при обработке информации о значениях некоторого показателя до и после воздействия на анализируемые объекты формулируется как:
 - Нет эффекта воздействия - средние значения генеральных совокупностей, из которых получены две анализируемые зависимые выборки, равны.**
 - Нет эффекта воздействия - дисперсии генеральных совокупностей, из которых получены две анализируемые зависимые выборки, равны.
 - Есть эффект воздействия - средние значения генеральных совокупностей, из которых получены две анализируемые зависимые выборки, не равны.
 - Есть эффект воздействия - дисперсии генеральных совокупностей, из которых получены две анализируемые зависимые выборки, не равны.
- Таблицы сопряженности применяются для анализа информации, представленной в:

- a. Интервальной шкале.
 - b. Категорированной (номинальной) шкале.**
 - c. Шкале отношений.
 - d. Интервальной шкале и шкале отношений.
3. Регрессионный анализ в процессе обработки информации применяется для:
- a. Выявления зависимости некоторого количественного показателя (отклика) от одного или нескольких показателей (предикторов).**
 - b. Выявления зависимости между качественными показателями.
 - c. Выявления причинно-следственных связей между качественными показателями.
 - d. Сокращения размерности пространства исследуемых показателей.
4. Коэффициент корреляции Спирмена в процессе обработки информации используется для:
- a. Выявления меры зависимости между двумя показателями, выборки значений которых предварительно проранжированы.**
 - b. Выявления зависимости между двумя номинальными показателями.
 - c. Выявления меры нелинейной зависимости между двумя показателями.
 - d. Оценки качества нелинейного уравнения регрессии.
5. Под статистической гипотезой в процессе обработки информации понимается:
- a. Предположение, связанное с вероятностным распределением некоторого показателя или показателей, которое исследователь собирается проверить в результате эксперимента.**
 - b. Предположение, которое исследователь собирается проверить в результате эксперимента.
 - c. Предположение о сценариях развития событий в будущем.
 - d. Предположение о значениях некоторой ненаблюдаемой переменной в прошлом.
6. Ошибкой первого рода при работе со статистической гипотезой в процессе обработки информации считается:
- a. Ошибка принять гипотезу, если на самом деле она не верна.
 - b. Ошибка отвергнуть гипотезу, если на самом деле она верна.**

- c. Ошибка использовать конкретный статистический критерий для проверки данной гипотезы.
 - d. Вычислительная ошибка при расчете статистики критерия.
7. Альтернативной гипотезой при проверке статистических гипотез в процессе обработки информации считается:
- a. Предположение, которое принимается, если отклоняется гипотеза H_0 .**
 - b. Предположение, которое оказывается однозначно истинным, если отвергается гипотеза H_0 .
 - c. Предположение, при котором вероятность попадания статистики критерия в критическую область равна 1.
 - d. Предположение, при котором вероятность попадания статистики критерия в критическую область больше 0,9.
8. Ранговый дисперсионный анализ Краскела-Уоллиса в процессе анализа информации используется для:
- a. Проверки влияния некоторого категоризованного показателя на некоторый признак, заданный в шкале отношений.**
 - b. Проверки влияния некоторого показателя, заданного в шкале отношений, на некоторый номинальный показатель.
 - c. Построение функции, описывающей влияния некоторого категоризованного показателя на некоторый признак, заданный в шкале отношений.
 - d. Построение функции, описывающей влияния некоторого показателя, заданного в шкале отношений, на некоторый номинальный показатель.
9. Дисперсионный анализ ANOVA в процессе анализа информации используется для оценки равенства:
- a. Средних значений для двух или более выборок, полученных из нормальных генеральных совокупностей.**
 - b. Дисперсий для двух или более выборок, полученных из нормальных генеральных совокупностей.
 - c. Средних значений только для двух выборок, полученных из нормальных генеральных совокупностей.
 - d. Средних значений для двух или более выборок, полученных из произвольных генеральных совокупностей.

10. Дисперсионный анализ ANOVA в процессе анализа информации можно использовать для сравнения:

a. Двух или более независимых выборок.

b. Двух или более зависимых выборок.

c. Только двух зависимых выборок.

d. Только двух независимых выборок.

11. Факторный анализ в процессе анализа информации позволяет:

a. Находить обобщенные, непосредственно не измеряемые показатели, которые определяют значения исходных измеряемых показателей.

b. Среди исходных показателей выделять показатели, имеющие высокую корреляцию с со всеми другими показателями.

c. Находить обобщенные показатели, такие, что все исходные показатели выражаются в виде конкретных функций от данных обобщенных показателей.

d. Находить один обобщенный, непосредственно не измеряемый показатель, который определяет значения исходных измеряемых показателей.

12. Под временным рядом в процессе анализа информации понимается:

a. Ряд значений некоторого статистического показателя, взятых в хронологическом порядке времени.

b. Ряд значений некоторого показателя для совокупности объектов для одного момента времени.

c. Функция от времени, описывающая значения некоторого детерминированного показателя.

d. Функция от времени, описывающая значения некоторого стохастического показателя.

13. Модель экспоненциального сглаживания для временных рядов в процессе анализа информации используется для:

a. Прогнозирования будущих значений временного ряда.

b. Выделения сезонной компоненты временного ряда.

c. Выделения тренда временного ряда.

d. Нахождения периода временного ряда.

14. Решение задачи дискриминации в рамках дискриминантного анализа при проведении прикладных исследований можно применять для:

- A. Определения по каким показателям существенно различаются выделенные в рамках исследования классы.**
- B. Разбиения исследуемых объектов на кластеры, близкие по своим характеристикам.
- C. Классификации принадлежности объекта к определенным в рамках исследования классам.
- D. Прогнозирования значений важных с точки зрения исследования показателей в рамках выделенных классов

15. Решение задачи кластеризации в рамках исследования выделенной совокупности объектов можно применять для:

- A. Определения наиболее важных показателей с точки зрения поиска различий между объектами.
- B. Разбиения совокупности объектов на кластеры, близкие по своим характеристикам.**
- C. Классификации принадлежности объектов к выделенным классам.
- D. Прогнозирования изменения кластеров во времени.

16. Для исследования зависимости между двумя категоризованными показателями можно использовать:

- A. Т-тест для зависимых выборок.
- B. Т-тест для независимых выборок.
- C. Метод «таблицы сопряженности».**
- D. Регрессионный анализ.

17. Для анализа зависимости между порядковыми показателями можно использовать:

- A. Т-тест для зависимых выборок.
- B. Т-тест для независимых выборок.
- C. Ранговый коэффициент корреляции Спирмена.**
- D. Регрессионный анализ.

18. Для снижения размерности признакового пространства можно использовать:

- A. Регрессионный анализ.
- B. Кластерный анализ.
- C. Факторный анализ.**
- D. Дисперсионный анализ

19. Автокорреляционная функция временного ряда для каждого значения аргумента, равного натуральному числу g , показывает:

- A. Как коррелируют наблюдения временного ряда, расположенные через g лагов**
- B. Как коррелируют соседние наблюдения g и $g+1$.
- C. Чему равна дисперсия g -го значения ряда.
- D. Чему равно математическое ожидание g -го значения ряда.

20. Для исследования значимых изменений в поведении некоторого показателя после воздействия на анализируемые объекты можно использовать:

- A. Т-тест для зависимых выборок.**
- B. Т-тест для независимых выборок.
- C. Ранговый коэффициент корреляции Спирмена.**
- D. Регрессионный анализ.

21. Для графического анализа гипотезы о нормальном распределении некоторого показателя можно использовать:

- A. Нормальную вероятностную бумагу.**
- B. Граф k -средних.
- C. Граф «ящики с усами».**
- D. Граф рассеивания.

22. Граф k -средних в кластерном анализе отражает:

- A. Центроиды кластеров.**
- B. Расстояние между центроидами кластеров.
- C. Разброс вокруг центроидов кластеров.
- D. Расстояние между ближайшими соседями кластеров.

23. Метод древовидной кластеризации используется:

A. Для анализа близости между объектами.

B. Для определения центроидов кластеров.

C. Для вычисления матрицы расстояний между кластерами.

D. Для определения показателей, которые нужно использовать при кластеризации.

Вопросы с кратким ответом

24. Компонента временного ряда, описывающая долговременную тенденцию его развития называется:

a. трендом.

25. При анализе информации о временных параметрах игроков некоторой игры за 2015-2016 годы выдвигается гипотеза:

H0: Среднее время игроков не изменилось.

H1: Среднее время игроков изменилось.

Результаты t-теста для зависимых выборок приведены в таблице:
Определите, отвергается гипотеза H0 (да или нет):

| T-test for Dependent Samples (Sheet1 in Book2) | | | | | | | | | | |
|--|----------|----------|----|----------|---------------|----------|----|----------|---------------------|---------------------|
| Marked differences are significant at $p < ,05000$ | | | | | | | | | | |
| Variable | Mean | Std.Dv. | N | Diff. | Std.Dv. Diff. | t | df | p | Confidence -95,000% | Confidence +95,000% |
| Минуты(2015г) | 1180,487 | 835,0211 | | | | | | | | |
| Минуты(2016г) | 1338,553 | 856,4092 | 76 | -158,066 | 364,3299 | -3,78225 | 75 | 0,000310 | -241,319 | -74,8128 |

a. Да.

26. Применение факторного анализа при обработке многомерной информации направлено на *увеличение* или *сокращение* размерности признакового пространства (укажите: увеличение или сокращение)

a Сокращение

27. В методах кластеризации многомерных данных в качестве меры близости объектов используется метрика или коэффициент корреляции (укажите: метрика или коэффициент корреляции)

a. метрика

28. Т-тест для зависимых выборок, одна из которых отражает значения некоторого показателя до реализации стратегии управления, а другая – после реализации

стратегии управления, используется для оценки равенства *средних* значений или *медиан* данных выборок (укажите: средних или медиан):

А. Средних

29. Метод сглаживания скользящими средними в анализе временных рядов динамики развития некоторого показателя используется для *выделения тренда* или *стандартизации ряда* (укажите: выделения тренда или стандартизации ряда)

А. Выделения тренда

30. Метод сезонной декомпозиции в анализе временных рядов динамики развития некоторого показателя используется для *исследования структуры* или *прогнозирования* (укажите: исследования структуры или прогнозирования)

А. Исследования структуры

31. Метод дискриминантного анализа предполагает, что объекты уже разбиты на классы (укажите: да или нет)

А. Да

Критерии и шкалы оценивания заданий ФОС:

Для оценивания выполнения заданий используется балльная шкала:

1) закрытые задания (тестовые с вариантами ответов, средний уровень сложности):

- 1 балл – указан верный ответ;
- 0 баллов – указан неверный ответ (полностью или частично неверный).

2) открытые задания (тестовые с кратким текстовым ответом, повышенный уровень сложности):

- 2 балла – указан верный ответ;
- 0 баллов – указан неверный ответ (полностью или частично неверный).

Задания раздела 20.3 рекомендуются к использованию при проведении диагностических работ с целью оценки остаточных результатов освоения данной дисциплины (знаний, умений, навыков).