

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
**«Российский государственный гуманитарный университет»**  
**(ФГБОУ ВО «РГУ»)**

ИНСТИТУТ ИНФОРМАЦИОННЫХ НАУК И ТЕХНОЛОГИЙ БЕЗОПАСНОСТИ  
Факультет информационных систем и безопасности  
Кафедра фундаментальной и прикладной математики

**АНАЛИЗ ДАННЫХ В СОЦИОТЕХНИЧЕСКИХ СИСТЕМАХ**  
**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

01.04.04 Прикладная математика

---

*Код и наименование направления подготовки/специальности*

**Математические методы и модели обработки  
и защиты информации в социотехнических системах**

---

*Наименование направленности (профиля)/ специализации*

Уровень высшего образования: *магистратура*

Форма обучения: *Очная, заочная*

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

Москва 2026

## АНАЛИЗ ДАННЫХ В СОЦИОТЕХНИЧЕСКИХ СИСТЕМАХ

Рабочая программа дисциплины

Составитель:

кандидат физ.-мат. наук, доц., доцент кафедры фундаментальной и прикладной математики  
*Синицын В.Ю.*

УТВЕРЖДЕНО

Протокол заседания кафедры  
фундаментальной и прикладной математики  
№ 5 от 19.12.2025

## ОГЛАВЛЕНИЕ

|      |   |    |
|------|---|----|
| 1.   | <a href="#">Пояснительная записка</a> .....   | 4  |
| 1.1. | <a href="#">Цель и задачи дисциплины</a> .....  | 4  |
| 1.2. | <a href="#">Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций</a> .....           | 4  |
| 1.3. | <a href="#">Место дисциплины в структуре образовательной программы</a> .....  | 4  |
| 2.   | <a href="#">Структура дисциплины</a> .....  | 5  |
| 3.   | <a href="#">Содержание дисциплины</a> .....   | 6  |
| 4.   | <a href="#">Образовательные технологии</a> .....  | 7  |
| 5.   | <a href="#">Оценка планируемых результатов обучения</a> .....   | 7  |
| 5.1. | <a href="#">Система оценивания</a> .....  | 7  |
| 5.2. | <a href="#">Критерии выставления оценки по дисциплине</a> .....   | 8  |
| 5.3. | <a href="#">Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине</a> ..... | 9  |
| 6.   | <a href="#">Учебно-методическое и информационное обеспечение дисциплины</a> .....   | 15 |
| 6.1. | <a href="#">Список источников и литературы</a> .....  | 15 |
| 6.2. | <a href="#">Перечень ресурсов информационно-телекоммуникационной сети «Интернет»</a> .....  | 15 |
| 6.3. | <a href="#">Профессиональные базы данных и информационно-справочные системы</a> .....   | 16 |
| 7.   | <a href="#">Материально-техническое обеспечение дисциплины</a> .....  | 16 |
| 8.   | <a href="#">Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов</a> .....                      | 17 |
| 9.   | <a href="#">Методические материалы</a> .....  | 18 |
| 9.1. | <a href="#">Планы практических занятий</a> .....  | 18 |
| 9.2. | <a href="#">Методические рекомендации по подготовке письменных работ</a> .....  | 26 |
|      | <a href="#">Приложение 1. Аннотация рабочей программы дисциплины</a> .....  | 27 |

## 1. Пояснительная записка

### 1.1. Цель и задачи дисциплины

*Цель дисциплины:* формирование у студентов современных представлений об анализе данных в социотехнических системах с использованием реальных данных и актуальных прикладных задач, а также о содержании и перспективах развития новой научной отрасли Big Data.

*Задачи дисциплины:* познакомить студентов с современными алгоритмами и технологиями автоматического быстрого анализа больших объёмов разнородной информации в социотехнических системах, развивать у студентов практические навыки анализа данных и интерпретации результатов исследования для решения прикладных задач.

### 1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

| Компетенция<br>(код и наименование)  | Индикаторы компетенций<br>(код и наименование)  | Результаты обучения  |
|--|---|--|
| ПК-1. Способен проводить систематизацию, алгоритмизацию конкретных информационных потоков по месту научных исследований, производственной деятельности | ПК-1.1.<br>Переформулирует задачи, данные на естественных языках конкретного научного знания на необходимый язык математики; формулирует теоремы. | <p><i>Знать:</i> основные стандартные типы прикладных задач, решаемых при помощи обработки данных и машинного обучения — классификация, регрессия, кластеризация, методы машинного обучения и их особенности, методы оценивания качества моделей, современные библиотеки для работы с моделями и оценки их качества</p> <p><i>Уметь:</i> работать с большими объемами данных, структурировать их, согласно требованиям заказчика, а также проводить анализ моделей различных типов, применять различные методы анализа данных для решения прикладных задач в социотехнических системах, разрабатывать и исследовать математические модели объектов, систем, процессов и технологий, предназначенных для проведения расчетов, анализа, подготовки решений, проводить научные эксперименты, оценивать результаты исследований</p> <p><i>Владеть:</i> навыками постановки прикладных задач, выбора соответствующих методов для их решения, анализа полученных результатов, а также навыками построения моделей и модификации стандартных методов при решении прикладных задач</p> |

### 1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Анализ данных в социотехнических системах» относится к части, формируемой участниками образовательных отношений, блока дисциплин учебного плана.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: «Математические методы исследования социальных систем», «Иностранный язык в профессиональной деятельности», «Принципы построения математических моделей в социотехнических системах», «Методология исследовательской деятельности и академическая культура», «Математические модели в истории науки и техники».

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: «Программные средства научного исследования», «Искусственные нейронные сети и интеллектуальный анализ данных», «Конструктивная математика и ее приложения в моделировании сложных систем», «Интеллектуальные системы», «Современные системы программирования», научно-исследовательская работа.

## 2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 5 з.е., 180 академических часов.

### Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

| Семестр | Тип учебных занятий  | Количество часов |
|---------|----------------------|------------------|
| 2       | Лекции               | 16               |
| 2       | Практические занятия | 34               |
| Всего:  |                      | 50               |

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 130 академических часа(ов).

### Структура дисциплины для заочной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

| Семестр | Тип учебных занятий  | Количество часов |
|---------|----------------------|------------------|
| 2       | Лекции               | 8                |
| 2       | Практические занятия | 12               |
| Всего:  |                      | 20               |

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 160 академических часов.

## 3. Содержание дисциплины

### Тема 1. Современные программные средства для статистического анализа данных.

Классификация задач прикладной статистики и методов их решения. Выбор соответствующего задаче метода обработки данных. Виды статистических пакетов. Требования к статистическим пакетам общего назначения. Сравнительный анализ возможностей по обработке данных, которые предоставляют системы Statistica, SPSS, R, SAS, Stata, Minitab, Statgraphics, Microsoft Excel Analysis ToolPak, STADIA и др. Особенности российского рынка программного обеспечения прикладной статистики.

### Тема 2. Анализ данных при помощи пакета Statistica.

Архитектура пакета Statistica. Интерфейс пользователя. Управление данными. Дизайн и сопровождение статистических баз данных. Встроенный язык программирования STATISTICA Visual Basic. Использование внешних языков программирования. Сетевые возможности пакета.

Обзор статистических методов, реализованных в пакете. Графические инструменты анализа данных. Технология обработки данных и подготовки отчетов. Многомерные статистические методы в пакете Statistica. Корреляционный и регрессионный анализ, анализ таблиц сопряженности, кластерный и дискриминантный анализ, факторный анализ, дисперсионный анализ, многомерное шкалирование и анализ надёжности, методы статистического контроля качества продукции, анализ выживаемости. Анализ временных рядов и прогнозирование. Моделирование структурными уравнениями (модуль SEPATH). Нейросетевой пакет STATISTICA Neural Networks и его использование для решения прикладных задач.

### **Тема 3. Анализ данных при помощи пакета SPSS.**

Архитектура пакета SPSS. Специфика оконного интерфейса. Редактор данных и вывод результатов вычислений. Дизайн и сопровождение статистических баз данных. Собственные средства программирования системы SPSS. Пакеты «R Essentials» и «Python Essentials». Интеграция SPSS с другими средствами статистических вычислений и языками программирования. Сетевые возможности SPSS. Обзор статистических методов, реализованных в пакете. Графические инструменты анализа данных. Технология обработки данных и подготовки отчетов. Многомерные статистические методы в пакете SPSS. Корреляционный и регрессионный анализ, анализ таблиц сопряженности, кластерный и дискриминантный анализ, факторный анализ, дисперсионный анализ, многомерное шкалирование и анализ надёжности, методы статистического контроля качества продукции, анализ выживаемости. Анализ временных рядов и прогнозирование. Моделирование структурными уравнениями. Нейросетевой модуль Neural Networks и его использование для решения прикладных задач.

### **Тема 4. Вычислительная среда R и ее использование для анализа данных.**

Исторические сведения о среде статистических вычислений и языке программирования R. Установка R в различных операционных системах. Режим командной строки, скрипты, базовые и рекомендованные пакеты. Сообщество разработчиков, техническая поддержка пользователей, документация, книги, журналы, регулярные международные конференции по языку R и его приложениям. Типы данных в R и принципы работы с ними. Числовые векторы, факторы, пропущенные данные, матрицы, списки. Таблицы данных. Векторизованные вычисления. Графические средства языка R. Два типа графических команд. Графические устройства и графические опции. Сохранение результатов работы. Статистическая обработка данных. Описательная статистика. Одномерные статистические тесты. Создание своих функций. Параметрические и непараметрические критерии проверки однородности выборок. Проверка гипотез нормальности распределения. Корреляционный анализ и анализ таблиц сопряженности. Графические интерфейсы к R: RStudio, RCommander, RKWard, JGR, SciViews-K, Rattle, PMG, RPMG, RWeb, gnumeric, Emacs и др. Поддержка работы с языком R в текстовых редакторах и средах разработки. Интеграция R с системами SPSS и Statistica. Интеллектуальный анализ данных (Data Mining) с помощью R. Графический анализ многих переменных.

### **Тема 5. Анализ данных при помощи универсальных математических пакетов.**

Статистические средства и инструментальные средства разработки универсальных математических пакетов MathCAD, Mathematica, MatLab, Maple, Maxima и др. Создание пользовательских интерфейсов к вычислительным процедурам.

### **Тема 6. Анализ данных при помощи офисных пакетов.**

Microsoft Excel Analysis ToolPak. Установка пакета и специфика интерфейса. Использование диалоговых окон. Подготовка данных. Создание, редактирование и печать диаграмм. Инструментарий статистического анализа данных и его использование. Описательная статистика. Генераторы случайных чисел. Создание выборки. Корреляции и ковариации. Двухвыборочный F-тест. T-тест двухвыборочный с одинаковыми и неодинаковыми дисперсиями. T-тест парный двухвыборочный для средних. Однофакторный и двухфакторный

дисперсионный анализ с повторением и без повторения. Экспоненциальное сглаживание. Скользящее среднее. Парная и множественная линейная регрессия. Другие статистические средства офисных пакетов.

#### 4. Образовательные технологии

Для проведения *занятий лекционного типа* по дисциплине применяются такие образовательные технологии как лекция-визуализация с применением слайд-проектора, лекция с разбором конкретных ситуаций, проблемная лекция.

Для проведения *практических занятий* используются такие образовательные технологии как: решение типовых задач для закрепления и формирования знаний, умений, навыков.

В рамках *самостоятельной работы* студентов проводится консультирование и проверка домашних заданий посредством электронной почты.

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

#### 5. Оценка планируемых результатов обучения

##### 5.1 Система оценивания

| Форма контроля                     | Макс. количество баллов |                   |
|------------------------------------|-------------------------|-------------------|
|                                    | За одну работу          | Всего             |
| Текущий контроль:                  |                         |                   |
| - опрос                            | 2 балла                 | 10 баллов         |
| - тестирование                     | 10 баллов               | 10 баллов         |
| - расчётно-графическая работа      | 20 баллов               | 20 баллов         |
| - доклад                           | 20 баллов               | 20 баллов         |
| Промежуточная аттестация – экзамен |                         |                   |
| - ответы на вопросы билета         | 10 баллов               | 20 баллов         |
| - итоговая контрольная работа      | 20 баллов               | 20 баллов         |
| <b>Итого за семестр</b>            |                         | <b>100 баллов</b> |

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

| 100-балльная шкала | Традиционная шкала  |            | Шкала ECTS |
|--------------------|---------------------|------------|------------|
| 95 – 100           | отлично             | зачтено    | A          |
| 83 – 94            |                     |            | B          |
| 68 – 82            | хорошо              |            | C          |
| 56 – 67            | удовлетворительно   |            | D          |
| 50 – 55            |                     |            | E          |
| 20 – 49            | неудовлетворительно | не зачтено | FX         |

|        |  |   |
|--------|--|---|
| 0 – 19 |  | F |
|--------|--|---|

## 5.2 Критерии выставления оценки по дисциплине

| Баллы/<br>Шкала<br>ECTS | Оценка по<br>дисциплине | Критерии оценки результатов обучения по дисциплине  |
|-------------------------|-------------------------|---|
| 100-83/<br>A,B          | отлично                 | <p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>                                   |
| 82-68/<br>C             | хорошо                  | <p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>                                       |
| 67-50/<br>D,E           | удовлетворительно       | <p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p> |
| 49-0/<br>F,FX           | неудовлетворительно     | <p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>               |

## 5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

### Текущий контроль

#### *Примерные задания для тестирования*

**по теме «Анализ данных в пакете Statistica»:**

Решите задачи, используя систему Statistica и файл с данными, который содержит результаты социологического опроса и личностные психологические показатели студентов РГГУ.

**Задача 1.**

Для девушек, степень религиозности которых сильная, среднее значение переменной E1\_Доброжелательность (с точностью до 0,01) равно

Ответ \_\_\_\_\_

**Задача 2.**

С помощью критерия Стьюдента (Т-критерия) выясните, на каком уровне значимости (с точностью до 0,001) различаются генеральные средние показателя N3\_Депрессивность для юношей и девушек.

Ответ \_\_\_\_\_

**Задача 3.**

С помощью критерия Колмогорова-Смирнова выясните, какие из приведенных ниже психологических показателей статистически значимо различаются для студентов факультета информатики (ФИ) и историко-филологического факультета (ИФФ).

Ответ 1. N3\_Депрессивность

Ответ 2. N4\_Застенчивость

Ответ 3. E3\_Настойчивость

Ответ 4. E4\_Активность

Ответ 5. O3\_Чувства

Ответ 6. O4\_Действия

Ответ 7. A3\_Альтруизм

Ответ 8. A4\_Уступчивость

Ответ 9. C3\_Ответственность

Ответ 10. C4\_Целеустремленность

**Задача 4.**

Коэффициент корреляции Спирмена пунктов I31 и I61 опросника NEO PI-R (с точностью до 0,001) равен

Ответ \_\_\_\_\_

**Задача 5.**

Для респондентов юношей постройте линейную регрессионную модель для психологического показателя N2\_Враждебность методом пошагового исключения независимых переменных, в качестве которых рассматривайте все остальные подшкалы теста NEO PI-R.

Коэффициент детерминации для полученной оптимальной модели с точностью до 0,001 равен

Ответ \_\_\_\_\_

**Задача 6.**

С помощью кластерного анализа методом К средних классифицируйте юношей с низким личным доходом на четыре класса, используя утверждения теста NEO PI-R от I21 до I120

Для полученной классификации расстояние от респондента с номером 176 до центра кластера, в котором он находится, (с точностью до 0,001) равно

Ответ \_\_\_\_\_

**Задача 7.**

Постройте наилучшую теоретическую классификацию респондентов на две группы, соответствующих степени согласия с утверждением I22 : 2 - “не согласен”, 4 - “согласен”. При построении классификации используйте метод пошагового дискриминантного анализа с включением независимых переменных, в качестве которых рассматривайте все тридцать подшкал теста NEO PI-R.

Для построенной классификации процент правильно теоретически распознанных ответов респондентов 2 - “не согласен” с точностью до 0,1% равен

Ответ \_\_\_\_\_

### Задача 8.

Для респондентов юношей с помощью критерия Шапиро-Уилка выясните, какие из приВыполните факторный анализ для респондентов девушек, используя данные по всем тридцати подшкалам теста NEO PI-R. Для выделения факторов примените метод Главных компонент с последующим Варимакс вращением. Классифицируйте подшкалы теста NEO PI-R, включив каждую из них в свою группу, соответствующую фактору, с которым у этой подшкалы наибольший (по абсолютной величине) коэффициент корреляции.

Используя построенную классификацию, укажите шкалы теста NEO PI-R из приведенного ниже списка, которые не пригодны для интерпретации фактора 2

Ответ 1. N\_Нейротизм

Ответ 2. E\_Экстраверсия

Ответ 3. O\_Открытость опыту

Ответ 4. A\_Согласие

Ответ 5. C\_Сознательность

### Задача 9.

Психометрическая подшкала N3\_Депрессивность теста NEO PI-R равна сумме восьми переменных (пунктов подшкалы) inv\_111, l41, inv\_171, l101, l131, l161, l191, l221. Выполните анализ пригодности этой подшкалы.

Наибольший из коэффициентов корреляции подшкалы со своими пунктами с точностью до 0,001 равен

Ответ \_\_\_\_\_

### Задача 10.

С помощью многомерного шкалирования (процедура ALSCAL) постройте двумерную модель множества всех подшкал теста NEO PI-R, используя данные только для респондентов с 51 до 350. При этом учитывайте, что шкала измерения данных Интервальная, а расстояние вычисляйте по формуле Расстояние Евклида.

Из приведенных ниже психологических показателей укажите три подшкалы, которые в построенной модели находятся дальше остальных (из этого списка) от подшкалы O1\_Фантазия

Ответ 1. N1\_Тревожность

Ответ 2. N2\_Враждебность

Ответ 3. N3\_Депрессивность

Ответ 4. N4\_Застенчивость

Ответ 5. N5\_Импульсивность

Ответ 6. N6\_Уязвимость

Ответ 7. E1\_Доброжелательность

Ответ 8. E2\_Общительность

Ответ 9. E3\_Настойчивость

Ответ 10. E4\_Активность

**Примерные задания для расчетно-графической работы  
по теме «Анализ данных в вычислительной среде R»:**

Решите задачи, используя вычислительную среду R и файл с данными, который содержит результаты социологического опроса и личностные психологические показатели студентов РГГУ.

**Задача 1.**

Инициализировать датчик случайных чисел с номером 2017000 и сгенерировать выборку объёма  $n=230$  из генеральной совокупности, имеющей показательный закон распределения с параметром  $rate=0.4$ . Найти с точностью до 0.01 выборочную квантиль на уровне 0.95.

Ответ \_\_\_\_\_

**Задача 2.**

Загрузить в рабочее пространство системы R данные из файла "NEO", который содержит результаты социологического опроса и личностные психологические показатели студентов. Используя фрейм данных с именем "NEO", найти число юношей, для которых значение переменной N6\_Уязвимость больше 28.

Ответ \_\_\_\_\_

**Задача 3.**

Инициализировать датчик случайных чисел с номером 2017000 и сгенерировать выборку объёма  $n=230$  из генеральной совокупности, имеющей закон распределения Пуассона с параметром  $lambda=8.1$ . По полученной выборке найти методом моментов с точностью до 0.01 точечную оценку параметра  $lambda$ , используя центральный момент второго порядка.

Ответ \_\_\_\_\_

**Задача 4.**

Инициализировать датчик случайных чисел с номером 2017000 и сгенерировать выборку объёма  $n=300$  из генеральной совокупности, имеющей нормальный закон распределения с параметрами  $mean=172$ ,  $sd=6.4$ . По полученной выборке найти с надёжностью  $p=0.95$  бутстреп-оценку доверительного интервала для математического ожидания, используя в качестве точечной оценки среднее арифметическое. Вычисления выполнить на основе 10000 вторичных выборок с объёмом 300 элементов каждая. В ответе указать длину доверительного интервала с точностью до 0.01.

Ответ \_\_\_\_\_

**Задача 5.**

Загрузить в рабочее пространство системы R данные из файла "NEO", который содержит результаты социологического опроса и личностные психологические показатели студентов. Используя фрейм данных с именем "NEO", с помощью критерия Колмогорова-Смирнова проверить статистическую гипотезу о том, что для респондентов юношей переменная O1\_Фантазия имеет закон распределения, который статистически значимо не отличается от нормального закона распределения. В ответе задачи указать значение  $p$ -value с точностью до 0.001.

Ответ \_\_\_\_\_

**Задача 6.**

Загрузить в рабочее пространство системы R данные из файла "NEO", который содержит результаты социологического опроса и личностные психологические показатели студентов. Используя фрейм данных с именем "NEO", с помощью критерия Краскела-Уоллиса проверить статистическую гипотезу о том, что уровень переменной O1\_Фантазия не зависит от семейного дохода респондентов. В ответе задачи указать значение  $p$ -value с точностью до 0.001.

Ответ \_\_\_\_\_

**Задача 7.**

Загрузить в рабочее пространство системы R данные из файла “NEO”, который содержит результаты социологического опроса и личностные психологические показатели студентов. Используя фрейм данных с именем “NEO”, с помощью Хи-квадрат критерия Пирсона проверить статистическую гипотезу о том, что ответы респондентов на пункт П\_41 опросника NEO PI-R не зависят от степени религиозности. В ответе задачи указать значение p-value с точностью до 0.001.

Ответ \_\_\_\_\_

**Задача 8.**

Загрузить в рабочее пространство системы R данные из файла “NEO”, который содержит результаты социологического опроса и личностные психологические показатели студентов. Используя фрейм данных с именем “NEO”, выяснить на уровне значимости 0.05, какие из перечисленных ниже порядковых демографических переменных имеют статистически значимый коэффициент корреляции Кендалла с психологическим показателем E5\_Непоседливость.

Ответ 1. возраст

Ответ 2. обр\_род (образование родителей)

Ответ 3. степ\_рел (степень религиозности)

Ответ 4. сем\_дох (семейный доход)

Ответ 5. лич\_дох (личный доход)

**Задача 9.**

Загрузить в рабочее пространство системы R данные из файла “NEO”, который содержит результаты социологического опроса и личностные психологические показатели студентов. Используя фрейм данных с именем “NEO”, построить оптимальную линейную регрессионную модель m26, содержащую 8 предикторов и переменную отклика N6\_Уязвимость, пошаговым методом добавления независимых переменных, в качестве которых рассматривать все подшкалы теста NEO PI-R кроме показателя N6\_Уязвимость. Найти с точностью до 0.001 коэффициент детерминации модели m26.

Ответ \_\_\_\_\_

**Задача 10.**

Загрузить в рабочее пространство системы R данные из файла “NEO”, который содержит результаты социологического опроса и личностные психологические показатели студентов. Используя фрейм данных с именем “NEO”, построить оптимальную линейную регрессионную модель m25 для психологического показателя N6\_Уязвимость пошаговым методом добавления независимых переменных, в качестве которых рассматривать все остальные подшкалы теста NEO PI-R. Используя модель m25, найти с точностью до 0.01 прогноз значения зависимой переменной N6\_Уязвимость для респондента с номером 208.

Ответ \_\_\_\_\_

***Примерные темы докладов***

1. Численный ресамплинг и его реализация в среде R.
2. Разработка интерфейсов для вычислительной среды R.
3. Бутстреп-оценки параметров распределений и их свойства.
4. Статистическое моделирование в вычислительной среде R.
5. Экспериментальное исследование мощности некоторых статистических критериев.
6. Исследование статистической устойчивости пятифакторной модели личности.
7. Статистические методы построения новых психометрических шкал.

8. Структурное моделирование психологического портрета личности с помощью теста NEO PI-R.
9. Характеристики качества датчиков псевдослучайных чисел.

*Примерные вопросы для опроса см. п.9.1 Планы практических занятий, контрольные вопросы*

**Промежуточная аттестация (экзамен)**  
**Примерные контрольные вопросы по курсу:**

1. Классификация задач прикладной статистики и методов их решения.
2. Виды статистических пакетов. Требования к статистическим пакетам.
3. Архитектура пакета Statistica. Интерфейс пользователя. Управление данными.
4. Встроенный язык программирования STATISTICA Visual Basic.
5. Многомерные статистические методы в пакете Statistica.
6. Моделирование структурными уравнениями (модуль SEPATH).
7. Анализ временных рядов и прогнозирование в системе Statistica.
8. Нейросетевой пакет STATISTICA Neural Networks и его применение.
9. Архитектура пакета SPSS. Специфика оконного интерфейса. Редактор данных и вывод результатов вычислений.
10. Собственные средства программирования системы SPSS. Интеграция SPSS с другими средствами статистических вычислений и языками программирования.
11. Многомерные статистические методы в пакете SPSS.
12. Анализ временных рядов и прогнозирование в системе SPSS.
13. Моделирование структурными уравнениями в SPSS.
14. Нейросетевой модуль Neural Networks и его применение
15. Общие сведения о среде статистических вычислений и языке программирования R.
16. Типы данных в R: векторы, факторы, пропущенные данные, матрицы, списки. Таблицы данных. Векторизованные вычисления.
17. Графические средства языка R. Два типа графических команд. Графические устройства и графические опции.
18. Статистическая обработка данных в системе R. Описательная статистика.
19. Проверка статистических гипотез в системе R.
20. Корреляционный анализ и анализ таблиц сопряженности. Регрессионный анализ в системе R.
21. Интеллектуальный анализ данных (Data Mining) с помощью R. Графический анализ многих переменных.
22. Статистические средства универсальных математических пакетов.
23. Microsoft Excel Analysis ToolPak. Другие статистические средства офисных пакетов.

**Примерные практические задания:**

Для решения задач использовать фрейм данных NEO из файла с именем "NEO", который содержит результаты социологического опроса и личностные психологические показатели студентов.

**Задача 1.**

С помощью критерия Шапиро-Уилка проверить статистическую гипотезу о том, что для респондентов девушек переменная N1\_Тревожность имеет закон распределения, который статистически значимо не отличается от нормального закона распределения. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 2.**

С помощью критерия Колмогорова-Смирнова проверить статистическую гипотезу о том, что для респондентов юношей переменная N1\_Тревожность имеет закон распределения,

который статистически значимо не отличается от нормального закона распределения. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 3.**

С помощью критерия Стьюдента проверить гипотезу о том, что математическое ожидание переменной N1\_Тревожность для студентов факультета «А» равно 27. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 4.**

С помощью критерия Стьюдента проверить гипотезу о том, что для юношей и девушек математические ожидания переменной N1\_Тревожность равны. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 5.**

С помощью критерия Уилкоксона проверить гипотезу о том, что положение переменной N1\_Тревожность для студентов факультета «А» равно 27. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 6.**

С помощью критерия Уилкоксона проверить гипотезу о том, что для респондентов с сильной и слабой степенью религиозности уровни переменной N1\_Тревожность равны. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 7.**

С помощью теста пропорций проверить статистическую гипотезу о том, что для студентов факультета «А» и студентов факультета «Б» доли респондентов, согласных с утверждением П\_31 опросника NEO PI-R, равны. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 8.**

С помощью Хи-квадрат критерия Пирсона проверить статистическую гипотезу о том, что ответы респондентов на пункт П\_31 опросника NEO PI-R не зависят от степени религиозности. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 9.**

С помощью однофакторного дисперсионного анализа проверить статистическую гипотезу о том, что математическое ожидание переменной N1\_Тревожность не зависит от семейного дохода респондентов. В ответе задачи указать значение p-value с точностью до 0.001.

**Задача 10.**

С помощью критерия Краскела-Уоллиса проверить статистическую гипотезу о том, что уровень переменной N1\_Тревожность не зависит от семейного дохода респондентов. В ответе задачи указать значение p-value с точностью до 0.001.

## 6. Учебно-методическое и информационное обеспечение дисциплины

### 6.1 Список источников и литературы

#### Литература

##### Основная

1. Груздев, А.В. Прогнозное моделирование в IBM SPSS Statistics и R: Метод деревьев решений / А.В. Груздев. - Москва : ДМК Пресс, 2016. - 278 с.
2. Наследов А. SPSS 19: профессиональный статистический анализ данных. - СПб.: Питер, 2011. - 400 с.
3. Основы эконометрики в пакете STATISTICA: Учебное пособие / Плохотников К.Э. - М.:Вузовский учебник, 2018. - 298 с. (Переплёт) ISBN 978-5-9558-0114-8 - Режим доступа: <http://znanium.com/catalog/product/914118>
4. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2024. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/536007> .
5. Миркин, Б. Г. Базовые методы анализа данных : учебник и практикум для вузов / Б. Г. Миркин. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2024. — 303 с. — (Высшее образование). — ISBN 978-5-534-18842-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/551786> .
6. Попова, И. Н. Анализ временных рядов : учебник для вузов / И. Н. Попова ; ответственный редактор В. В. Ковалев. — Москва : Издательство Юрайт, 2024. — 74 с. — (Высшее образование). — ISBN 978-5-534-18394-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/534918>.

##### Дополнительная

1. Дятлов, А.В. Анализ данных в социологии : учебник / А.В.Дятлов, Д.А.Гугуева ; Южный федеральный университет. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2018. - 226 с. - ISBN 978-5-9275-2690-1. - Режим доступа: <https://new.znanium.com/catalog/product/1039664>
2. Маккинли, У. Python и анализ данных / Уэс Маккинли ; пер. с англ. А.А. Слинкина. - Москва : ДМК Пресс, 2015. - 482 с.
3. Тюрин Ю. Н. Анализ данных на компьютере: учеб. пособие по направлениям "Математика", "Математика. Прикладная математика" / Ю. Н. Тюрин, А. А. Макаров. - Изд. 4-е, перераб. - М.: Форум, 2013. - 366 с.- (Высшее образование)
4. Храмов, Д.А. Сбор данных в Интернете на языке R / Д. А. Храмов. - Москва : ДМК Пресс, 2017. - 280 с.

### 6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

1. Учебно-образовательная физико-математическая библиотека на портале МИР МАТЕМАТИЧЕСКИХ УРАВНЕНИЙ [Электронный ресурс]. - Режим доступа: <http://eqworld.ipmnet.ru/ru/library.htm>
2. Аникин Ю. и др. Введение в аналитику больших массивов данных. Дистанционный учебный курс на портале Intuit.ru [Электронный ресурс]. - Режим доступа: <http://www.intuit.ru/studies/courses/12385/1181/info>
3. Мастицкий С.Э., Шитиков В.К. (2014) Статистический анализ и визуализация данных с помощью R. – Электронная книга, адрес доступа: <https://github.com/ranalytics/r-tutorials>
4. Чубукова И.А. Data Mining. Учеб. курс НОУ ИНТУИТ [Электронный ресурс]. - Режим доступа: <http://www.intuit.ru/departament/database/datamining/>

5. Шипунов А.Б., Балдин Е.М., Волкова П.А., Коробейников А.И., Назарова С.А., Петров С.В., Суфиянов В.Г. Наглядная статистика. Используем R! — М.: ДМК Пресс, 2012. - 298 с. [Электронный ресурс]. - Режим доступа: <http://herba.msu.ru/shipunov/software/r/r-ru.htm>

Национальная электронная библиотека (НЭБ) [www.rusneb.ru](http://www.rusneb.ru)  
 ELibrary.ru Научная электронная библиотека [www.elibrary.ru](http://www.elibrary.ru)

### 6.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

Информационные справочные системы:

1. Консультант Плюс
2. Гарант

### 7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения:

- для лекций: учебные аудитории, оснащённые доской, компьютером или ноутбуком, проектором (стационарным или переносным) для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows
2. Microsoft Office
3. Kaspersky Endpoint Security

- для практических занятий: компьютерный класс или лаборатория, оснащённые доской, компьютером или ноутбуком для преподавателя, компьютерами для обучающихся, проектором (стационарным или переносным) для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows
2. Microsoft Office
3. Mozilla Firefox
4. Язык программирования R
5. SPSS
6. Statistica
7. Kaspersky Endpoint Security

Для практических занятий можно также использовать актуальные полнофункциональные демонстрационные версии профессиональных статистических пакетов SPSS и Statistica.

### 8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением;

обеспечением или могут быть заменены устным ответом; обеспечивается индивидуальное равномерное освещение не менее 300 люкс; для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; письменные задания оформляются увеличенным шрифтом; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

- для глухих и слабослышащих: лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования; письменные задания выполняются на компьютере в письменной форме; экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих: в печатной форме увеличенным шрифтом, в форме электронного документа, в форме аудиофайла.

- для глухих и слабослышащих: в печатной форме, в форме электронного документа.

- для обучающихся с нарушениями опорно-двигательного аппарата: в печатной форме, в форме электронного документа, в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих: устройством для сканирования и чтения с камерой SARA CE; дисплеем Брайля PAC Mate 20; принтером Брайля EmBraille ViewPlus;

- для глухих и слабослышащих: автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих; акустический усилитель и колонки;

- для обучающихся с нарушениями опорно-двигательного аппарата: передвижными, регулируемые эргономическими партами СИ-1; компьютерной техникой со специальным программным обеспечением.

## **9. Методические материалы**

### **9.1 Планы практических занятий**

Форма проведения – решение типовых задач для закрепления и формирования знаний, умений, навыков

#### **Тема 1. Пакет Statistica.**

**Цель работы** - знакомство с технологией статистического анализа данных в пакете Statistica.

**Контрольные вопросы:**

1. Описательная статистика в пакете Statistica.
2. Проверка статистических гипотез в пакете Statistica.
3. Дисперсионный анализ в пакете Statistica.
4. Корреляционный анализ в пакете Statistica.
5. Множественная линейная регрессия в пакете Statistica.
6. Кластерный анализ в пакете Statistica.
7. Дискриминантный анализ в пакете Statistica.
8. Факторный анализ в пакете Statistica.
9. Анализ надежности в пакете Statistica.
10. Многомерное шкалирование в пакете Statistica.
11. Статистический анализ временных рядов в пакете Statistica.

**Примерные задачи для решения в аудитории:**

При решении задач рекомендуется использовать файл с данными, который содержит результаты социологического опроса и личностные психологические показатели студентов.

**Задача 1.**

Для девушек, степень религиозности которых слабая, среднее значение переменной E2\_Общительность (с точностью до 0,01) равно

Ответ 25,78

**Задача 2.**

С помощью критерия Стьюдента (Т-критерия) выясните, на каком уровне значимости (с точностью до 0,001) различаются генеральные средние показателя N2\_Враждебность для юношей и девушек.

Ответ 0,005

**Задача 3.**

С помощью критерия Манна-Уитни выясните, на каком уровне значимости различаются генеральные средние показателя E5\_Непоседливость для девушек с сильной и слабой степенью религиозности.

Ответ 0,024

**Задача 4.**

Коэффициент корреляции Спирмена пунктов I31 и I51 опросника NEO PI-R (с точностью до 0,001) равен

Ответ 0,184

**Задача 5.**

Для респондентов юношей постройте линейную регрессионную модель для психологического показателя N1\_Тревожность методом пошагового исключения независимых переменных, в качестве которых рассматривайте все остальные подшкалы теста NEO PI-R. Коэффициент детерминации для полученной оптимальной модели с точностью до 0,001 равен

Ответ 0,671

**Задача 6.**

С помощью кластерного анализа методом К средних классифицируйте юношей с низким личным доходом на четыре класса, используя утверждения теста NEO PI-R от I21 до I120. Для полученной классификации расстояние от респондента с номером 148 до центра кластера, в котором он находится, (с точностью до 0,001) равно

Ответ 0,705

**Задача 7.**

Для множества респондентов с 31 до 230 постройте наилучшую теоретическую классификацию студентов на две группы - "мужчины" и "женщины", используя метод пошагового дискриминантного анализа с включением независимых переменных, в качестве

которых рассматривайте все тридцать подшкал теста NEO PI-R. Для построенной классификации процент правильно теоретически распознанных респондентов девушек с точностью до 0,1% равен

Ответ 94,0

#### **Задача 8.**

Выполните факторный анализ для множества респондентов с 31 до 230, используя данные по всем тридцати подшкалам теста NEO PI-R. Для выделения факторов примените метод Главных компонент с последующим Варимакс вращением. Накопленный процент объясненной дисперсии данных для 5 извлеченных факторов с точностью до 0,001 равен

Ответ 59,228%

#### **Задача 9.**

Психометрическая подшкала N4\_Застенчивость теста NEO PI-R равна сумме восьми переменных (пунктов подшкалы) 116, inv\_146, 176, inv\_1106, 1136, inv\_1166, 1196, inv\_1226. Выполните анализ пригодности этой подшкалы. Показатель надёжности альфа Кронбаха для этой подшкалы с точностью до 0,001 равен

Ответ 0,753

#### **Задача 10.**

С помощью многомерного шкалирования (процедура ALSCAL) постройте двумерную модель множества всех подшкал теста NEO PI-R, используя данные только для респондентов с 51 до 350. При этом учитывайте, что шкала измерения данных Интервальная, а расстояние вычисляйте по формуле Расстояние Евклида. В построенной модели расстояние в двухмерном пространстве от подшкалы O1\_Фантазия до ближайшей к ней подшкалы с точностью до 0,001 равно

Ответ 0,257

### **Тема 2. Пакет SPSS.**

**Цель работы** - знакомство с технологией статистического анализа данных в пакете SPSS.

#### **Контрольные вопросы:**

1. Описательная статистика в пакете SPSS.
2. Проверка статистических гипотез в пакете SPSS.
3. Дисперсионный анализ в пакете SPSS.
4. Корреляционный анализ в пакете SPSS.
5. Множественная линейная регрессия в пакете SPSS.
6. Кластерный анализ в пакете SPSS.
7. Дискриминантный анализ в пакете SPSS.
8. Факторный анализ в пакете SPSS.
9. Анализ надёжности в пакете SPSS.
10. Многомерное шкалирование в пакете SPSS.
11. Статистический анализ временных рядов в пакете SPSS.

#### **Примерные задачи для решения в аудитории:**

При решении задач рекомендуется использовать файл с данными, который содержит результаты социологического опроса и личностные психологические показатели студентов.

#### **Задача 1.**

Число респондентов, семейный доход которых низкий, равно

Ответ 22

#### **Задача 2.**

С помощью критерия Стьюдента (Т-критерия) выясните, какие из приведенных ниже психологических показателей статистически значимо различаются для юношей и девушек.

- Ответ 1. +N1\_Тревожность
- Ответ 2. +N2\_Враждебность
- Ответ 3. E1\_Доброжелательность
- Ответ 4. E2\_Общительность
- Ответ 5. +O1\_Фантазия
- Ответ 6. +O2\_Эстетичность
- Ответ 7. A1\_Доверие
- Ответ 8. A2\_Прямота
- Ответ 9. +C1\_Компетентность
- Ответ 10. +C2\_Организованность

**Задача 3.**

С помощью критерия Колмогорова-Смирнова выясните, какие из приведенных ниже психологических показателей статистически значимо различаются для студентов факультета информатики (ФИ) и историко-филологического факультета (ИФФ).

- Ответ 1. +N1\_Тревожность
- Ответ 2. +N2\_Враждебность
- Ответ 3. E1\_Доброжелательность
- Ответ 4. E2\_Общительность
- Ответ 5. +O1\_Фантазия
- Ответ 6. +O2\_Эстетичность
- Ответ 7. A1\_Доверие
- Ответ 8. A2\_Прямота
- Ответ 9. C1\_Компетентность
- Ответ 10. +C2\_Организованность

**Задача 4.**

Выясните, какие из перечисленных ниже порядковых демографических переменных имеют статистически значимый коэффициент корреляции Спирмена с психологическим показателем E2\_Общительность.

- Ответ 1. +возраст
- Ответ 2. +обр\_род (образование родителей)
- Ответ 3. степ\_рел (степень религиозности)
- Ответ 4. сем\_дох (семейный доход)
- Ответ 5. +лич\_дох (личный доход)

**Задача 5.**

Постройте линейную регрессионную модель для психологического показателя C6\_Осмотрительность методом пошагового включения независимых переменных, в качестве которых рассматривайте все остальные подшкалы теста NEO PI-R.

Коэффициент детерминации для модели, содержащей 7 самых важных независимых переменных, с точностью до 0,001 равен

Ответ 0,448

**Задача 6.**

С помощью иерархического кластерного анализа классифицируйте тридцать подшкал теста NEO PI-R на пять классов, используя данные только для множества респондентов с 51 до 350. В качестве метода кластеризации примените метод Внутригрупповые связи, а расстояние вычисляйте по формуле Расстояние Евклида.

По результатам классификации выясните, какие из приведенных ниже психологических показателей относятся к кластеру 2

- Ответ 1. N3\_Депрессивность
- Ответ 2. N4\_Застенчивость
- Ответ 3. E3\_Настойчивость
- Ответ 4. E4\_Активность
- Ответ 5. +O3\_Чувства

- Ответ 6. O4\_Действия
- Ответ 7. +A3\_Альтруизм
- Ответ 8. A4\_Уступчивость
- Ответ 9. C3\_Ответственность
- Ответ 10. C4\_Целеустремленность

**Задача 7.**

Для множества респондентов с 51 до 350 постройте наилучшую теоретическую классификацию студентов на две группы - “мужчины” и “женщины”, используя метод пошагового дискриминантного анализа с включением и исключением независимых переменных, в качестве которых рассматривайте все тридцать подшкал теста NEO PI-R. При вычислении учитывайте относительные размеры групп.

Используя построенную классификацию, укажите номера респондентов из приведенного ниже списка, для которых принадлежность к группе распознана неверно

- Ответ 1. +92
- Ответ 2. 93
- Ответ 3. 94
- Ответ 4. 95
- Ответ 5. 96
- Ответ 6. +97
- Ответ 7. 98
- Ответ 8. +99
- Ответ 9. 100
- Ответ 10. +101

**Задача 8.**

Выполните факторный анализ для множества респондентов с 51 до 350, используя данные по всем тридцати подшкалам теста NEO PI-R. Для выделения факторов примените метод Главных компонент с последующим Варимакс вращением. Классифицируйте подшкалы теста NEO PI-R, включив каждую из них в свою группу, соответствующую фактору, с которым у этой подшкалы наибольший (по абсолютной величине) коэффициент корреляции.

Используя построенную классификацию, укажите подшкалы из приведенного ниже списка, которые включены в группу, соответствующую фактору 2

- Ответ 1. +A1\_Доверие
- Ответ 2. +A2\_Прямота
- Ответ 3. +A3\_Альтруизм
- Ответ 4. A4\_Уступчивость
- Ответ 5. A5\_Скромность
- Ответ 6. +A6\_Отзывчивость
- Ответ 7. C1\_Компетентность
- Ответ 8. C2\_Организованность
- Ответ 9. C3\_Ответственность
- Ответ 10. C4\_Целеустремленность

**Задача 9.**

Психометрическая подшкала A3\_Альтруизм теста NEO PI-R равна сумме восьми переменных (пунктов подшкалы) inv\_114, l44, inv\_174, l104, inv\_1134, l164, l194, l224. Выполните анализ пригодности этой подшкалы.

Показатель надёжности альфа Кронбаха для этой подшкалы с точностью до 0,001 равен

- Ответ 0,658

**Задача 10.**

С помощью многомерного шкалирования (процедура ALSCAL) постройте двумерную модель множества всех подшкал теста NEO PI-R, используя данные только для респондентов с 51 до 350. При этом учитывайте, что шкала измерения данных Интервальная, а расстояние вычисляйте по формуле Расстояние Евклида.

Из приведенных ниже психологических показателей укажите три подшкалы, которые в построенной модели находятся дальше остальных (из этого списка) от подшкалы O1\_Фантазия

- Ответ 1. N1\_Тревожность
- Ответ 2. +N2\_Враждебность
- Ответ 3. N3\_Депрессивность
- Ответ 4. N4\_Застенчивость
- Ответ 5. N5\_Импульсивность
- Ответ 6. +N6\_Уязвимость
- Ответ 7. E1\_Доброжелательность
- Ответ 8. E2\_Общительность
- Ответ 9. +E3\_Настойчивость
- Ответ 10. E4\_Активность

### **Тема 3. Вычислительная среда и язык программирования R.**

**Цель работы** - знакомство с технологией статистического анализа данных в среде R.

#### **Контрольные вопросы:**

1. Описательная статистика в среде R.
2. Графические методы анализа данных в среде R.
3. Проверка статистических гипотез в среде R.
4. Дисперсионный анализ в среде R.
5. Корреляционный анализ в среде R.
6. Регрессионный анализ в среде R.

#### **Примерные задачи для решения в аудитории:**

При решении задач рекомендуется использовать файл с данными, который содержит результаты социологического опроса и личностные психологические показатели студентов.

#### **Задача 1.**

Загрузить в рабочее пространство системы R данные из файла “NEO”, который содержит результаты социологического опроса и личностные психологические показатели студентов. Используя фрейм данных с именем “NEO”, выполнить следующие задания: 1) вывести на экран имена всех переменных и найти количество строк и столбцов таблицы данных “NEO”; 2) выяснить структуру данных части таблицы “NEO”, содержащей первые пять столбцов; 3) получить сводную информацию о переменных с номерами 3, 12 и 17; 4) записать сводную информацию обо всех переменных таблицы “NEO” в текстовый файл “NEO.summary.txt”; 5) создать подмножество фрейма данных “NEO”, которое содержит информацию о студентках факультета “Б” с сильной степенью религиозности, и найти количество строк полученного фрейма данных; 6) графически исследовать степень религиозности респондентов (переменная “СТЕП\_РЕЛ”); 7) построить график переменной N6\_Уязвимость; 8) построить график зависимости личного дохода от пола респондента; 9) построить график зависимости показателя N6\_Уязвимость от пола респондента; 10) построить график зависимости показателя N2\_Враждебность от показателя N6\_Уязвимость.

#### **Задача 2.**

Используя фрейм данных с именем “NEO”, выполнить следующие задания: 1) построить таблицу частот и таблицу относительных частот для переменной СТЕП\_РЕЛ; 2) составить список, компонентами которого являются таблицы частот для переменных с номерами от 2 до 7 (включительно); 3) построить таблицу частот для переменной N6\_Уязвимость; 4) для группированного вариационного ряда переменной N6\_Уязвимость построить таблицы частот, относительных частот, накопленных и накопленных относительных частот.

**Задача 3.**

Используя фрейм данных с именем “NEO”, выполнить следующие задания: 1) построить полигон относительных частот для переменной N6\_Уязвимость и добавить на график кривую плотности распределения вероятностей нормального закона, параметрами которого считать выборочное среднее и выборочное стандартное отклонение переменной N6\_Уязвимость; 2) построить полигон накопленных относительных частот для переменной N6\_Уязвимость и добавить к нему график функции распределения нормального закона с такими же параметрами, как в задании 1; 3) выполнить задание 1 для группированного вариационного ряда переменной N6\_Уязвимость с 10 интервалами группировки; 4) выполнить задание 2 для группированного вариационного ряда переменной N6\_Уязвимость с 10 интервалами группировки.

**Задача 4.**

Используя фрейм данных с именем “NEO”, найти эмпирическую функцию распределения для переменной N6\_Уязвимость. Построить её график. С помощью эмпирической функции распределения вычислить статистические вероятности следующих событий: А — переменная N6\_Уязвимость принимает значение не больше 20; В — переменная N6\_Уязвимость принимает значение на отрезке [25, 30]; С — переменная N6\_Уязвимость принимает значение больше 35.

**Задача 5.**

Используя фрейм данных с именем “NEO”, найти для переменной N6\_Уязвимость следующие числовые характеристики выборки: объём  $n$ ; наименьшее и наибольшее значения  $\min$  и  $\max$ ; выборочное среднее (арифметическое)  $m$ ; медиану  $m_e$ ; нижнюю и верхнюю квартили  $q_1$  и  $q_3$ ; (исправленную) выборочную дисперсию  $\text{var}$ ; среднее квадратическое отклонение  $\text{sd}$ ; размах выборки  $g$ ; межквартильный размах  $\text{iqr}$ ; моду  $m_0$ , начальные  $\alpha[i]$  и центральные  $\text{mu}[i]$  моменты до 4 порядка включительно ( $i=1,2,3,4$ ), асимметрию  $\text{as}$ , эксцесс  $\text{ex}$ .

**Задача 6.**

Используя фрейм данных с именем “NEO”, выполнить следующие задания: 1) для девушек факультета «Б», которые младше 19 лет, найти наибольшее значение переменной N6\_Уязвимость; 2) для юношей со слабой степенью религиозности найти с точностью до 0.001 среднее арифметическое переменной N6\_Уязвимость; 3) указать знак зодиака, для которого выборочная дисперсия показателя N6\_Уязвимость является наибольшей; 4) выбрать знаки зодиака, для которых выборочное среднее показателя N6\_Уязвимость меньше 22.

**Задача 7.**

Используя фрейм данных с именем “NEO”, построить с помощью функции `boxplot()` следующие графики: 1) диаграмму распределения переменных N6\_Уязвимость, E6\_Жизнерадостность, A6\_Отзывчивость, C6\_Осмотрительность; 2) диаграмму зависимости распределения переменной N6\_Уязвимость от уровня фактора ФАКУЛЬТ.

**Задача 8.**

Используя фрейм данных с именем “NEO”, найти с надёжностью  $p=0.95$  интервальную бутстреп-оценку `bcir1` для коэффициента корреляции Пирсона случайных величин N6\_Уязвимость и C6\_Осмотрительность. Вычисление интервальной бутстреп-оценки выполнить на основе 10000 вторичных выборок с объёмом, равным объёму исходной выборки.

**Задача 9.**

Инициализировать датчик случайных чисел с номером 2017000 и сгенерировать выборку объёма  $n=300$  из генеральной совокупности, имеющей нормальный закон распределения с параметрами  $\text{mean}=172$ ,  $\text{sd}=6.4$ . По полученной выборке найти с надёжностью  $p=0.95$  интервальные бутстреп-оценки квантилей на уровнях 0.05, 0.25, 0.5, 0.75, 0.95. Вычисление интервальных бутстреп-оценок выполнить на основе 10000 вторичных выборок с объёмом 300 элементов каждая. Построить график зависимости интервальной бутстреп-оценки верхней квартили от доверительной вероятности.

**Задача 10.**

Инициализировать датчик случайных чисел с номером 2017000 и сгенерировать выборку объёма  $n=300$  из генеральной совокупности, имеющей нормальный закон распределения с параметрами  $\text{mean}=172$ ,  $\text{sd}=6.4$ . По полученной выборке найти с надёжностью  $p=0.95$  интервальную бутстреп-оценку асимметрии генеральной совокупности. Вычисление интервальной бутстреп-оценки выполнить на основе 10000 вторичных выборок с объёмом 300 элементов каждая. Построить график зависимости этой интервальной бутстреп-оценки от доверительной вероятности.

**Задача 11.**

Используя фрейм данных с именем “NEO”, который содержит результаты социологического опроса и личностные психологические показатели студентов, с помощью критерия Шапиро-Уилка проверить следующие статистические гипотезы: 1) переменная N6\_Уязвимость имеет закон распределения, который статистически значимо не отличается от нормального закона распределения; 2) для респондентов юношей переменная N6\_Уязвимость имеет закон распределения, который не отличается от нормального закона распределения.

**Задача 12.**

Используя фрейм данных с именем “NEO”, с помощью критерия Колмогорова-Смирнова проверить следующие статистические гипотезы: 1) переменная N6\_Уязвимость имеет нормальный закон распределения с параметрами, которые равны выборочному среднему и выборочному стандартному отклонению; 2) для респондентов юношей переменная N6\_Уязвимость имеет нормальный закон распределения с параметрами, которые равны выборочному среднему и выборочному стандартному отклонению.

**Задача 13.**

Используя фрейм данных с именем “NEO”, с помощью критерия Стьюдента проверить гипотезу о том, что для юношей и девушек математические ожидания переменной N6\_Уязвимость равны.

**Задача 14.**

Используя фрейм данных с именем “NEO”, с помощью критерия Уилкоксона проверить следующие статистические гипотезы: 1) уровень переменной N6\_Уязвимость равен 22; 2) уровень переменной N6\_Уязвимость для студентов факультета «А» равен 22.

**Задача 15.**

Используя фрейм данных с именем “NEO”, с помощью критерия Уилкоксона проверить следующие статистические гипотезы: 1) для юношей и девушек уровень различий переменной N6\_Уязвимость равен нулю; 2) для респондентов с сильной и слабой степенью религиозности уровень различий переменной N6\_Уязвимость равен нулю.

**Задача 16.**

Используя фрейм данных с именем “NEO”, проверить следующие статистические гипотезы: 1) коэффициент корреляции Пирсона переменных С6\_Осмотрительность и N6\_Уязвимость равен нулю; 2) коэффициент корреляции Кендалла переменных С6\_Осмотрительность и N6\_Уязвимость равен нулю; 3) коэффициент корреляции Спирмена переменных С6\_Осмотрительность и N6\_Уязвимость равен нулю.

**Задача 17.**

Используя фрейм данных с именем “NEO”, с помощью хи-квадрат критерия Пирсона проверить следующие статистические гипотезы: 1) степень религиозности не зависит от пола респондентов; 2) степень религиозности не зависит от семейного дохода; 3) переменные П\_3 и П\_12 независимы.

**Задача 18.**

Используя фрейм данных с именем “NEO”, с помощью однофакторного дисперсионного анализа проверить статистическую гипотезу о том, что математическое ожидание переменной N6\_Уязвимость не зависит от семейного дохода респондентов.

#### **Задача 19.**

Используя фрейм данных с именем “NEO”, с помощью критерия Краскела-Уоллиса проверить статистическую гипотезу о том, что уровень переменной N6\_Уязвимость не зависит от семейного дохода респондентов.

#### **Задача 20.**

Среди случайно взятых 10000 новорождённых оказалось 5143 мальчика. С помощью теста пропорций проверить статистическую гипотезу о том, что вероятность рождения мальчика равна 0.5.

#### **Задача 21.**

Используя фрейм данных с именем “NEO”, с помощью теста пропорций проверить следующие статистические гипотезы: 1) для юношей и девушек вероятности сильной степени религиозности равны; 2) для студентов разных факультетов вероятности сильной степени религиозности равны.

#### **Задача 22.**

Используя фрейм данных с именем “NEO”, с помощью критерия Бартлетта проверить следующие статистические гипотезы: 1) дисперсия переменной N6\_Уязвимость одинаковая для разных факультетов; 2) дисперсия переменной N6\_Уязвимость одинаковая для разных уровней семейного дохода респондентов; 3) дисперсия переменной N6\_Уязвимость одинаковая для разных знаков зодиака.

#### **Задача 23.**

Используя фрейм данных с именем “NEO”, найти выборочные коэффициенты корреляции Пирсона переменных С6\_Осмотрительность и N6\_Уязвимость для девушек и юношей отдельно. На уровне значимости 0.05 проверить являются ли эти коэффициенты корреляции статистически значимыми.

#### **Задача 24.**

Загрузить фрейм данных NEO. На уровне значимости 0.1 найти статистически значимые коэффициенты корреляции Спирмена показателя А6\_Отзывчивость с порядковыми демографическими переменными ВОЗРАСТ, СТЕП\_РЕЛ, СЕМ\_ДОХ, ЛИЧ\_ДОХ.

#### **Задача 25.**

Загрузить фрейм данных NEO. На уровне значимости 0.1 найти статистически значимые коэффициенты корреляции Кендалла переменной СТЕП\_РЕЛ со следующими показателями: N1\_Тревожность, N2\_Враждебность, E1\_Доброжелательность, E2\_Общительность, O1\_Фантазия, O2\_Эстетичность, A1\_Доверие, A2\_Прямота, C1\_Компетентность, C2\_Организованность.

#### **Задача 26.**

Загрузить фрейм данных NEO. Для студентов факультета “Б” с помощью критерия Фишера на уровне значимости 0.05 найти количество статистически значимых зависимостей среди первых тридцати пунктов опросника NEO PI-R. Решить ту же задачу с помощью хи-квадрат критерия Пирсона. Сравнить результаты вычислений.

#### **Задача 27.**

Загрузить фрейм данных NEO, который содержит результаты социологического опроса и личностные психологические показатели студентов. Построить линейную регрессионную модель m1 зависимости психологического показателя N6\_Уязвимость от показателей N1\_Тревожность, A1\_Доверие, O2\_Эстетичность. Найти основные характеристики модели m1 и построить модель m2, удалив из модели m1 независимую переменную A1\_Доверие. Оценить

характеристики модели m2 и построить модель m3, удалив из модели m2 независимую переменную O2\_Эстетичность. Сравнить качество моделей m2 и m3. Построить линейную регрессионную модель m4 зависимости психологического показателя N6\_Уязвимость от всех остальных 29 подшкал теста NEO PI-R. Оценить характеристики модели m4.

#### **Задача 28.**

Загрузить фрейм данных NEO. Построить линейную регрессионную модель m21 зависимости показателя N6\_Уязвимость от переменных N2\_Враждебность, O2\_Эстетичность, C2\_Организованность, A2\_Прямота, C5\_Самодисциплина. Оптимизировать модель m21 и построить модель m22, используя пошаговый метод автоматического исключения переменных на основе информационного критерия AIC. Сравнить качество моделей m21 и m22 по скорректированному коэффициенту детерминации.

#### **Задача 29.**

Загрузить фрейм данных NEO. Построить оптимальную линейную регрессионную модель m25 для психологического показателя N6\_Уязвимость пошаговым методом добавления независимых переменных, в качестве которых рассматривать все остальные подшкалы теста NEO PI-R. Найти число независимых переменных, включенных автоматически в модель m25.

#### **Задача 30.**

Загрузить фрейм данных NEO. Выполнить многофакторный дисперсионный анализ, построив следующие линейные модели: 1) двухфакторную модель m33 зависимости показателя цинизм от степени религиозности и пола респондентов; 2) модель m34 зависимости показателя цинизм от степени религиозности и семейного дохода респондентов с учётом взаимодействия этих факторов; 3) модель m35 зависимости показателя цинизм от степени религиозности, семейного дохода и факультета обучения с учётом всех взаимодействий между этими факторами; 4) модель m36, полученную из модели m35 пошаговым методом автоматического исключения независимых переменных.

## **9.2 Методические рекомендации по подготовке письменных работ**

Отчет по выполнению расчетно-графических работ по дисциплине «Анализ данных в социотехнических системах» объёмом 15-20 страниц выполняется студентом по каждой работе отдельно. Правила оформления отчета по выполнению расчетно-графических работ совпадают с правилами оформления курсовой работы, которые содержатся в «Методических рекомендациях по подготовке и оформлению курсовой работы» (официальный сайт кафедры ФПМ ИИНТБ РГГУ).

## АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина «Анализ данных в социотехнических системах» реализуется на факультете информационных систем и безопасности кафедрой фундаментальной и прикладной математики.

Цель дисциплины: формирование у студентов современных представлений об анализе данных в социотехнических системах с использованием реальных данных и актуальных прикладных задач, а также о содержании и перспективах развития новой научной отрасли Big Data.

Задачи: познакомить студентов с современными алгоритмами и технологиями автоматического быстрого анализа больших объёмов разнородной информации в социотехнических системах, развивать у студентов практические навыки анализа данных и интерпретации результатов исследования для решения прикладных задач.

Дисциплина направлена на формирование следующих компетенций:

- ПК-1. Способен проводить систематизацию, алгоритмизацию конкретных информационных потоков по месту научных исследований, производственной деятельности.

В результате освоения дисциплины обучающийся должен:

*Знать:* основные стандартные типы прикладных задач, решаемых при помощи обработки данных и машинного обучения — классификация, регрессия, кластеризация, методы машинного обучения и их особенности, методы оценивания качества моделей, современные библиотеки для работы с моделями и оценки их качества

*Уметь:* работать с большими объемами данных, структурировать их, согласно требованиям заказчика, а также проводить анализ моделей различных типов, применять различные методы анализа данных для решения прикладных задач в социотехнических системах, разрабатывать и исследовать математические модели объектов, систем, процессов и технологий, предназначенных для проведения расчетов, анализа, подготовки решений, проводить научные эксперименты, оценивать результаты исследований

*Владеть:* навыками постановки прикладных задач, выбора соответствующих методов для их решения, анализа полученных результатов, а также навыками построения моделей и модификации стандартных методов при решении прикладных задач.

По дисциплине предусмотрена промежуточная аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины составляет 5 зачетных единиц.