

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Луганский государственный университет имени Владимира Даля»

Экономический факультет  
Кафедра экономической кибернетики и прикладной статистики

УТВЕРЖДАЮ:  
Декан экономического факультета  
Тхор Е.С.  
(подпись) \_\_\_\_\_ 2023 года  
« 24 » \_\_\_\_\_



**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**

**«АНАЛИЗ ДАННЫХ СРЕДСТВАМИ  
ЯЗЫКА ПРОГРАММИРОВАНИЯ R»**

По направлению подготовки 38.03.05 Бизнес-информатика  
Профиль: «Экономическая аналитика и бизнес-статистика»

Луганск – 2023

Лист согласования РПУД


Рабочая программа учебной дисциплины «Анализ данных средствами языка программирования R» по направлению подготовки 38.03.05 Бизнес-информатика. – 30 с.

Рабочая программа учебной дисциплины «Анализ данных средствами языка программирования R» составлена с учетом Федерального государственного образовательного стандарта высшего образования по направлению подготовки 38.03.05 Бизнес-информатика, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 29 июня 2020 года № 838.

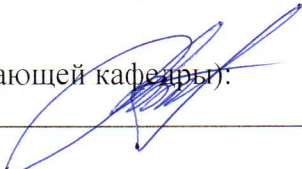
СОСТАВИТЕЛЬ (СОСТАВИТЕЛИ):

к.э.н., доцент Спорняк С.А.

Рабочая программа дисциплины утверждена на заседании кафедры экономической кибернетики и прикладной статистики «18» 04 2023 г., протокол № 26

Заведующий кафедрой экономической кибернетики и прикладной статистики  А.В. Велигура

Переутверждена: «  »    20   г., протокол №   

Согласована (для обеспечивающей кафедры):  
Декан экономического факультета  Тхор Е.С.

Переутверждена: «  »    20   года, протокол №   

Рекомендована на заседании учебно-методической комиссии экономического факультета «21» апреле 2023 г., протокол № 4.

Председатель учебно-методической комиссии экономического факультета  Е.Н. Шаповалова

## Структура и содержание дисциплины

### 1. Цели и задачи дисциплины, ее место в учебном процессе

Целью дисциплины является изучение методов, моделей и информационных технологий в области анализа данных. А также способов обработки массивов данных и проведения исследований в различных предметных областях.

Задачами изучения дисциплины являются:

изучение методов, моделей и средств анализа данных;

изучение методики проведения процедуры анализа данных;

применение на практике изученных методов и моделей (в качестве информационной технологии используется язык R).

### 2. Место дисциплины в структуре ОПОП ВО.

Дисциплина «Анализ данных средствами языка программирования R» относится к вариативной части (часть, формируемая участниками образовательных отношений) дисциплин.

Необходимыми условиями для освоения дисциплины являются: знания теории системного анализа, моделирования бизнес-процессов, архитектуры информационных и компьютерных систем.

Содержание дисциплины является логическим продолжением содержания дисциплин «Бизнес-информатика», «Бизнес-информатика 2», «Статистика», «Эконометрика» и служит основой для освоения дисциплин «Современные технологии обработки статистических данных», «Визуализация данных».

### 3. Требования к результатам освоения содержания дисциплины

| Код и наименование компетенции  | Индикаторы достижений компетенции (по реализуемой дисциплине)   | Перечень планируемых результатов  |
|---|---|---|
| ПК-6. Способен применять статистические методы для анализа экономических данных, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ | ПК-6.3. Использует прикладное современное программное обеспечение, умеет выбрать и применить оптимальную прикладную программу для решения конкретной задачи | <b>Знать:</b> прикладное современное программное обеспечение, применяемое при проведении исследований |
|   |   | <b>Уметь:</b> выбрать и применить оптимальную прикладную программу для решения конкретной задачи      |
|   |   | <b>Владеть:</b> навыками применения цифровых технологий для решения задач анализа данных              |

## 4. Структура и содержание дисциплины

### 4.1. Объем учебной дисциплины и виды учебной работы

| Вид учебной работы  | Объем часов (зач. ед.) |                    |                    |
|---|------------------------|--------------------|--------------------|
|   | Очная форма            | Очно-заочная форма | Заочная форма      |
| Общая учебная нагрузка (всего)  | 144<br>(4 зач. ед)     | 144<br>(4 зач. ед) | 144<br>(4 зач. ед) |
| Обязательная контактная работа (всего)<br>в том числе:  | 68                     | 28                 | 12                 |
| Лекции  | 34                     | 14                 | 6                  |
| Семинарские занятия   | -                      | -                  | -                  |
| Практические занятия  | -                      | -                  | -                  |
| Лабораторные работы   | 34                     | 14                 | 6                  |
| Курсовая работа (курсовой проект)   | -                      | -                  | -                  |
| Другие формы и методы организации образовательного процесса ( <i>расчетно-графические работы, индивидуальные задания и т.п.</i> ) | -                      | -                  | -                  |
| <b>Самостоятельная работа студента (всего)</b>  | <b>76</b>              | <b>116</b>         | <b>132</b>         |
| Форма аттестации  | зачет                  | зачет              | зачет              |

### 4.2. Содержание разделов дисциплины

#### Тема 1. ЗНАКОМСТВО С R

Зачем использовать R? Получение и установка R. Работа в R. Пакеты. Пакетная обработка. Использование вывода в качестве ввода – повторное использование результатов. Работа с большими массивами данных.

#### Тема 2. СОЗДАНИЕ НАБОРА ДАННЫХ

Что такое набор данных? Структуры данных. Ввод данных. Аннотирование наборов данных. Полезные функции для работы с объектами.

#### Тема 3. НАЧАЛО РАБОТЫ С ДИАГРАММАМИ

Работа с диаграммами. Графические параметры. Добавление текста, настройка параметров осей и условных обозначений. Объединение диаграмм.

#### Тема 4. ОСНОВЫ УПРАВЛЕНИЯ ДАННЫМИ

Создание новых переменных. Перекодировка переменных. Переименование переменных. Пропущенные значения. Календарные даты как данные. Преобразования данных из одного типа в другой. Сортировка данных. Объединение наборов данных. Разделение наборов данных на составляющие. Использование команд SQL для преобразования таблиц данных

#### Тема 5. БОЛЕЕ СЛОЖНЫЕ СПОСОБЫ УПРАВЛЕНИЯ ДАННЫМИ

Числовые и текстовые функции. Решение нашей задачи по управлению данными. Управление Агрегирование и изменение структуры данных выполнением команд. Функции, написанные пользователем.

#### Тема 6. БАЗОВЫЕ ДИАГРАММЫ

Столбчатые диаграммы. Круговые диаграммы. Гистограммы. Диаграммы ядерной оценки функции плотности. Диаграммы размахов. Точечные диаграммы.

## **Тема 7. ОСНОВНЫЕ МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ**

Описательные статистики. Таблицы частот и таблицы сопряженности. Корреляции. Тесты Стьюдента. Непараметрические тесты межгрупповых различий. Визуализация групповых различий.

### **Тема 8. РЕГРЕССИЯ**

МНК-регрессия. Диагностика регрессионных моделей. Необычные наблюдения. Способы корректировки. Выбор «лучшей» регрессионной модели. Продолжение анализа.

### **Тема 9. ДИСПЕРСИОННЫЙ АНАЛИЗ**

Подгонка ANOVA-моделей. Однофакторный дисперсионный анализ. Однофакторный ковариационный анализ. Двухфакторный дисперсионный анализ. Дисперсионный анализ для повторных измерений. Многомерный дисперсионный анализ. Дисперсионный анализ как регрессия.

### **Тема 10. АНАЛИЗ МОЩНОСТИ**

Проведение анализа мощности при помощи пакета `power`. Графический анализ мощности. Другие пакеты.

### **Тема 11. ДИАГРАММЫ СРЕДНЕЙ СЛОЖНОСТИ**

Диаграммы рассеяния. Линейные графики. Кореллограммы. Мозаичные диаграммы.

### **Тема 12. СТАТИСТИКА ПОВТОРНЫХ ВЫБОРОК И БУТСТРЕП-АНАЛИЗ**

Перестановочные тесты. Перестановочные тесты в пакете `coin`. Перестановочные тесты, реализованные в пакете `lmPerm`. Дополнительные замечания о перестановочных тестах. Бутстреп-анализ. Бутстреп-анализ при помощи пакета `boot`.

### **Тема 13. ОБОБЩЕННЫЕ ЛИНЕЙНЫЕ МОДЕЛИ**

Обобщенные линейные модели и функция `glm()`. Логистическая регрессия. Пуассоновская регрессия.

### **Тема 14. ГЛАВНЫЕ КОМПОНЕНТЫ И ФАКТОРНЫЙ АНАЛИЗ**

Выполнение анализа главных компонент и факторного анализа в R. Главные компоненты. Разведочный факторный анализ. Другие модели для латентных переменных.

### **Тема 15. ПРОДВИНУТЫЕ МЕТОДЫ РАБОТЫ С ПРОПУЩЕННЫМИ ДАННЫМИ**

Этапы работы с пропущенными данными. Обнаружение пропущенных значений. Исследование структуры пропущенных данных. Выявление источников пропущенных данных и эффекта от них. Рациональный подход. Анализ полных строк (построчное удаление). Метод множественного восстановления пропущенных данных. Другие подходы к пропущенным данным.

### **Тема 16. ПРОДВИНУТЫЕ ГРАФИЧЕСКИЕ МЕТОДЫ**

Четыре графические системы R. Пакет `lattice`. Пакет `ggplot2`. Интерактивная графика.

### 4.3. Лекции

| № п/п         | Название темы                                    | Объем часов |                    |               |
|---------------|--|-------------|--------------------|---------------|
|               |  | Очная форма | Очно-заочная форма | Заочная форма |
| 1.            | Знакомство с R                                   | 2           | 0,5                | 0,25          |
| 2.            | Создание набора данных                           | 2           | 0,5                | 0,25          |
| 3.            | Начало работы с диаграммами                      | 2           | 0,5                | 0,25          |
| 4.            | Основы управления данными                        | 2           | 0,5                | 0,25          |
| 5.            | Более сложные способы управления данными         | 2           | 1                  | 0,5           |
| 6.            | Базовые диаграммы                                | 2           | 1                  | 0,25          |
| 7.            | Основные методы статистической обработки данных  | 2           | 1                  | 0,25          |
| 8.            | Регрессия  | 2           | 1                  | 0,25          |
| 9.            | Дисперсионный анализ                             | 3           | 1                  | 0,5           |
| 10.           | Анализ мощности                                  | 2           | 1                  | 0,5           |
| 11.           | Диаграммы средней сложности                      | 2           | 1                  | 0,5           |
| 12.           | Статистика повторных выборок и бутстреп-анализ   | 3           | 1                  | 0,5           |
| 13.           | Обобщенные линейные модели                       | 2           | 1                  | 0,5           |
| 14.           | Главные компоненты и факторный анализ            | 2           | 1                  | 0,5           |
| 15.           | Продвинутые методы работы с пропущенными данными | 2           | 1                  | 0,5           |
| 16.           | Продвинутые графические методы                   | 2           | 1                  | 0,25          |
| <b>Итого:</b> |  | <b>34</b>   | <b>14</b>          | <b>6</b>      |

**4.4. Практические (семинарские) занятия по дисциплине «Анализ данных средствами языка программирования R» не предполагаются учебным планом.**

### 4.5. Лабораторные работы

| № п/п | Название темы                                   | Объем часов |                    |               |
|-------|---|-------------|--------------------|---------------|
|       |   | Очная форма | Очно-заочная форма | Заочная форма |
| 1.    | Знакомство с R                                  | 2           | 0,5                | 0,25          |
| 2.    | Создание набора данных                          | 2           | 0,5                | 0,25          |
| 3.    | Начало работы с диаграммами                     | 2           | 0,5                | 0,25          |
| 4.    | Основы управления данными                       | 2           | 0,5                | 0,25          |
| 5.    | Более сложные способы управления данными        | 2           | 1                  | 0,5           |
| 6.    | Базовые диаграммы                               | 2           | 1                  | 0,25          |
| 7.    | Основные методы статистической обработки данных | 2           | 1                  | 0,25          |
| 8.    | Регрессия                                       | 2           | 1                  | 0,25          |
| 9.    | Дисперсионный анализ                            | 3           | 1                  | 0,5           |
| 10.   | Анализ мощности                                 | 2           | 1                  | 0,5           |
| 11.   | Диаграммы средней сложности                     | 2           | 1                  | 0,5           |
| 12.   | Статистика повторных выборок и бутстреп-анализ  | 3           | 1                  | 0,5           |
| 13.   | Обобщенные линейные модели                      | 2           | 1                  | 0,5           |

|               |  |           |           |          |
|---------------|--|-----------|-----------|----------|
| 14.           | Главные компоненты и факторный анализ            | 2         | 1         | 0,5      |
| 15.           | Продвинутые методы работы с пропущенными данными | 2         | 1         | 0,5      |
| 16.           | Продвинутые графические методы                   | 2         | 1         | 0,25     |
| <b>Итого:</b> |  | <b>34</b> | <b>14</b> | <b>6</b> |

#### 4.6. Самостоятельная работа студентов

| № п/п | Название темы                                   | Вид СРС   | Объем часов |                    |               |
|-------|---|---|-------------|--------------------|---------------|
|       |   |   | Очная форма | Очно-заочная форма | Заочная форма |
| 1.    | Знакомство с R                                  | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 4           | 7                  | 8             |
| 2.    | Создание набора данных                          | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 4           | 7                  | 8             |
| 3.    | Начало работы с диаграммами                     | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 4           | 7                  | 8             |
| 4.    | Основы управления данными                       | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 4           | 7                  | 8             |
| 5.    | Более сложные способы управления данными        | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 4           | 7                  | 8             |
| 6.    | Базовые диаграммы                               | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 4           | 7                  | 8             |
| 7.    | Основные методы статистической обработки данных | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 4           | 7                  | 8             |
| 8.    | Регрессия                                       | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 4           | 7                  | 8             |
| 9.    | Дисперсионный анализ                            | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 5           | 7                  | 8             |
| 10.   | Анализ мощности                                 | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 5           | 7                  | 8             |
| 11.   | Диаграммы средней сложности                     | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 5           | 7                  | 8             |

|               |  |   |           |            |            |
|---------------|--|---|-----------|------------|------------|
| 12.           | Статистика повторных выборок и бутстреп-анализ   | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 5         | 7          | 8          |
| 13.           | Обобщенные линейные модели                       | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 5         | 7          | 8          |
| 14.           | Главные компоненты и факторный анализ            | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 5         | 7          | 8          |
| 15.           | Продвинутые методы работы с пропущенными данными | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 5         | 7          | 8          |
| 16.           | Продвинутые графические методы                   | Подготовка к лабораторным работам, к текущему и промежуточному контролю знаний. | 5         | 7          | 8          |
| 17.           | Зачет  | Подготовка к зачету   | 4         | 4          | 4          |
| <b>Итого:</b> |  |   | <b>76</b> | <b>116</b> | <b>132</b> |

**4.7. Курсовые работы/проекты по дисциплине «Анализ данных средствами языка программирования R» не предполагаются учебным планом.**

### **5. Образовательные технологии**

Преподавание дисциплины ведется с применением следующих видов образовательных технологий: объяснительно-иллюстративного обучения (технология поддерживающего обучения, технология проведения учебной дискуссии), информационных технологий (презентационные материалы), развивающих и инновационных образовательных технологий.

Лабораторные работы проводятся с использованием развивающих, проблемных, проектных, информационных (использование электронных образовательных ресурсов (электронный конспект) образовательных технологий.

### **6. Формы контроля освоения дисциплины**

Текущая аттестация студентов производится в дискретные временные интервалы лектором и преподавателем(ями), ведущими лабораторные работы по дисциплине в следующих формах:

тесты;

вопросы для обсуждения (в виде докладов и сообщений).

Промежуточная аттестации по результатам освоения дисциплины проходит в форме устного зачета.

В экзаменационную ведомость и зачетную книжку выставляются оценки по шкале, приведенной в таблице.



| Характеристика знания предмета и ответов  | Зачеты     |
|---|------------|
| Студент глубоко и в полном объёме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач. | зачтено    |
| Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.                                      |            |
| Студент знает только основной программный материал, допускает неточности, недостаточно чёткие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.                                    |            |
| Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.                           | не зачтено |

## 7. Учебно-методическое и информационное обеспечение дисциплины:

### а) основная литература:

1. Волкова В.М., Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python: учебное пособие / Волкова В.М. - Новосибирск: Изд-во НГТУ, 2017. - 74 с. - ISBN 978-5-7782-3183-2 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785778231832.htm>.
2. Интерактивный курс Introduction to R от Data Camp [Электронный ресурс] / Режим доступа - <https://www.datacamp.com/courses/free-introduction-to-r>
3. Интерактивный курс Try R [Электронный ресурс] / режим доступа - <http://tryr.codeschool.com/>.

### б) дополнительная литература:

1. Визуализация результатов ВНО 2016 в R. [Электронный ресурс] / Режим доступа - <http://rpubs.com/mgontar/214015>.
2. Соловьева С.В., Информационные технологии в профессиональной деятельности: инструментарий бизнес-аналитики: практикум [Электронный ресурс] / С.В. Соловьева, Ю.П. Александровская, Ю.В. Хайрутдинова - Казань: Издательство КНИТУ, 2017. - 104 с. - ISBN 978-5-7882-2217-2 // ЭБС "Консультант студента": <http://www.studentlibrary.ru/book/ISBN9785788222172.html>

3. BaseGroup Labs технологии анализа данных [Электронный ресурс] / Режим доступа - <http://basegroup.ru>.
4. Александровская Ю.П., Многомерный статистический анализ в экономике [Электронный ресурс]: учебное пособие / Александровская Ю. П. - Казань: Издательство КНИТУ, 2017. - 96 с. - ISBN 978-5-7882-2191-5 // ЭБС "Консультант студента": <http://www.studentlibrary.ru/book/ISBN9785788221915.htm>
5. Мельниченко А.С., Математическая статистика и анализ данных: учеб. пособие / А.С. Мельниченко - М.: МИСиС, 2018. - 45 с. - ISBN 978-5-906953-62-9 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785906953629.htm>
6. Храмов Д.А., Сбор данных в Интернете на языке R / Храмов Д. А. - М.: ДМК Пресс, 2017. - 280 с. - ISBN 978-5-97060-459-5 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785970604595.htm>
7. Коваленко Т.А., Обработка экспериментальных данных / Коваленко Т.А. - М.: Национальный Открытый Университет "ИНТУИТ", 2016. - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: [http://www.studentlibrary.ru/book/intuit\\_161.htm](http://www.studentlibrary.ru/book/intuit_161.htm).

#### **в) методические указания:**

1. Методические указания к лабораторной работе по дисциплине «Анализ данных средствами языка программирования R» для студентов направления подготовки «Бизнес-информатика». Тема: «Начало работы в RStudio» / сост. А.А. Лофиченко. – Луганск: ГОУ ВО ЛНР «ЛГУ им. В. Даля», 2020. – 38 с.
2. Методические указания к лабораторной работе по дисциплине «Анализ данных средствами языка программирования R» для студентов направления подготовки «Бизнес-информатика». Тема: «Создание набора данных» / сост. А.А. Лофиченко. – Луганск: ГОУ ВО ЛНР «ЛГУ им. В. Даля», 2020. – 34 с.
3. Методические указания к лабораторной работе по дисциплине «Анализ данных средствами языка программирования R» для студентов направления подготовки «Бизнес-информатика». Тема: «Начало работы с диаграммами» / сост. А.А. Лофиченко. – Луганск: ГОУ ВО ЛНР «ЛГУ им. В. Даля», 2020. – 40 с.
4. Методические указания к лабораторной работе по дисциплине «Анализ данных средствами языка программирования R» для студентов направления подготовки «Бизнес-информатика». Тема: «Основы управления данными» / сост. А.А. Бугаева, А.А. Лофиченко. – Луганск: ГОУ ВО ЛНР «ЛГУ им. В. Даля», 2021. – 29 с.
5. Методические указания к лабораторной работе по дисциплине «Анализ данных средствами языка программирования R» для студентов направления подготовки «Бизнес-информатика». Тема: «Более сложные

способы управления данными» / сост. А.А. Бугаева, А.А. Лофиченко. – Луганск: ГОУ ВО ЛНР «ЛГУ им. В. Даля», 2021. – 37 с.

6. Методические указания к лабораторной работе по дисциплине «Анализ данных средствами языка программирования R» для студентов направления подготовки «Бизнес-информатика». Тема: «Основные методы статистической обработки данных» / сост. А.А. Лофиченко. – Луганск: ЛГУ им. В. Даля, 2022. – 36 с.

7. Методические указания к лабораторной работе по дисциплине «Анализ данных средствами языка программирования R» для студентов направления подготовки «Бизнес-информатика». Тема: «Базовые диаграммы» / сост. А.А. Лофиченко. – Луганск: ЛГУ им. В. Даля, 2022. – 34 с.

8. Методические указания к лабораторной работе по дисциплине «Анализ данных средствами языка программирования R» для студентов направления подготовки «Бизнес-информатика». Тема: «Дисперсионный анализ» / сост. М.А. Пронина. – Луганск: ЛГУ им. В. Даля, 2023. – 38 с.

#### **г) Интернет-ресурсы:**

1. Министерство образования и науки Российской Федерации – <http://минобрнауки.рф/>

2. Федеральная служба по надзору в сфере образования и науки – <http://obrnadzor.gov.ru/>

3. Министерство образования и науки Луганской Народной Республики – <https://minobr.su>

4. Министерство промышленности и торговли Луганской Народной Республики – <https://www.minpromlnr.su/main.php/>

5. Министерство экономического развития Луганской Народной Республики – <https://merlnr.su/>

6. Министерство финансов Луганской Народной Республики – <https://minfinlnr.su/>

7. Народный совет Луганской Народной Республики – <https://nslnr.su>

8. Государственный комитет статистики Луганской Народной Республики – <https://www.gkslnr.su/>

9. Портал Федеральных государственных образовательных стандартов высшего образования – <http://fgosvo.ru/>

10. Федеральный портал «Российское образование» – <http://www.edu.ru/>

11. Информационная система «Единое окно доступа к образовательным ресурсам» – <http://window.edu.ru/>

12. Федеральный центр информационно-образовательных ресурсов – <http://fcior.edu.ru/>

#### **Электронные библиотечные системы и ресурсы**

13. Электронно-библиотечная система «Консультант студента» – <http://www.studentlibrary.ru/>

14. Электронно-библиотечная система «StudMed.ru» –  
<https://www.studmed.ru/>

**Информационный ресурс библиотеки образовательной организации**

15. Научная библиотека имени А. Н. Коняева –  
<http://biblio.dahluniver.ru/>.

## **8. Материально-техническое и программное обеспечение дисциплины**

Освоение дисциплины «Анализ данных средствами языка программирования R» предполагает использование академических аудиторий, соответствующих действующим санитарным и противопожарным правилам и нормам.

Прочее: рабочее место преподавателя, оснащенное компьютером с доступом в Интернет.

Программное обеспечение:

| <b>Функциональное назначение</b> | <b>Бесплатное программное обеспечение</b> | <b>Ссылки</b>   |
|----------------------------------|---|---|
| Офисный пакет                    | OpenOffice 4.3.7                          | <a href="https://www.openoffice.org/">https://www.openoffice.org/</a>                                       |
| Операционная система             | UBUNTU 19.04                              | <a href="https://ubuntu.com/">https://ubuntu.com/</a>   |
| Браузер                          | Firefox Mozilla                           | <a href="http://www.mozilla.org/ru/firefox/fx">http://www.mozilla.org/ru/firefox/fx</a>                     |
| Архиватор                        | 7Zip                                      | <a href="http://www.7-zip.org/">http://www.7-zip.org/</a>   |
| Графический редактор             | GIMP (GNU Image Manipulation Program)     | <a href="http://www.gimp.org/">http://www.gimp.org/</a>   |
| Редактор PDF                     | Adobe Acrobat Reader                      | <a href="https://get.adobe.com/ru/reader/">https://get.adobe.com/ru/reader/</a>                             |
| Аудиоплеер                       | VLC                                       | <a href="http://www.videolan.org/vlc/">http://www.videolan.org/vlc/</a>                                     |
| Язык программирования            | R   | <a href="https://cran.r-project.org/mirrors.html">https://cran.r-project.org/mirrors.html</a>               |
| Среда разработки                 | RStudio                                   | <a href="https://rstudio.com/products/rstudio/download/">https://rstudio.com/products/rstudio/download/</a> |

## 9. Оценочные средства по дисциплине

### Паспорт оценочных средств по учебной дисциплине

«Анализ данных средствами языка программирования R»

Перечень компетенций (элементов компетенций), формируемых в результате освоения учебной дисциплины (модуля) или практики

| № п/п | Код контролируемой компетенции | Формулировка контролируемой компетенции   | Индикаторы достижений компетенции (по реализуемой дисциплине) | Контролируемые темы учебной дисциплины, практики | Этапы формирования (семестр изучения) |
|-------|--------------------------------|---|---|--|---------------------------------------|
| 1     | ПК-6                           | Способен применять статистические методы для анализа экономических данных, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ | ПК-6.3  | Тема 1   | 5                                     |
|       |                                |   |   | Тема 2   |                                       |
|       |                                |   |   | Тема 3   |                                       |
|       |                                |   |   | Тема 4   |                                       |
|       |                                |   |   | Тема 5   |                                       |
|       |                                |   |   | Тема 6   |                                       |
|       |                                |   |   | Тема 7   |                                       |
|       |                                |   |   | Тема 8   |                                       |
|       |                                |   |   | Тема 9   |                                       |
|       |                                |   |   | Тема 10  |                                       |
|       |                                |   |   | Тема 11  |                                       |
|       |                                |   |   | Тема 12  |                                       |
|       |                                |   |   | Тема 13  |                                       |
|       |                                |   |   | Тема 14  |                                       |
|       |                                |   |   | Тема 15  |                                       |
|       |                                |   |   | Тема 16  |                                       |

### Показатели и критерии оценивания компетенций, описание шкал оценивания

| № п/п | Код контролируемой компетенции | Индикаторы достижений компетенции (по реализуемой дисциплине) | Перечень планируемых результатов   | Контролируемые темы учебной дисциплины | Наименование оценочного средства                            |
|-------|--------------------------------|---|--|--|---|
| 1.    | ПК-6                           | ПК-6.3  | знать прикладное современное программное обеспечение, применяемое при проведении исследований; | Тема 1                                 | Тесты, вопросы для обсуждения (в виде докладов и сообщений) |
|       |                                |   |  | Тема 2                                 |   |
|       |                                |   |  | Тема 3                                 |   |
|       |                                |   |  | Тема 4                                 |   |
|       |                                |   |  | Тема 5                                 |   |
|       |                                |   |  | Тема 6                                 |   |
|       |                                |   |  | Тема 7                                 |   |
|       |                                |   |  | Тема 8                                 |   |
|       |                                |   |  | Тема 9                                 |   |
|       |                                |   |  | Тема 10                                |   |

|  |  |  |         |
|--|--|--|---------|
|  |  | уметь выбрать и применить оптимальную прикладную программу для решения конкретной задачи; владеть навыками применения цифровых технологий для решения задач анализа данных | Тема 11 |
|  |  |  | Тема 12 |
|  |  |  | Тема 13 |
|  |  |  | Тема 14 |
|  |  |  | Тема 15 |
|  |  |  | Тема 16 |

## Фонды оценочных средств по дисциплине «Анализ данных средствами языка программирования R»

### Тесты

1. Концепция векторизации подразумевает, что:
  - а) некоторые функции и операторы действуют поэлементно
  - б) при создании вектора не нужно указывать его тип
  - в) вместо поэлементного присваивания можно воспользоваться функцией `s`
  - г) для операций с векторами необходимо использовать циклы в явном виде
  - д) индексация векторов начинается с единицы
  
2. Пусть целочисленная матрица `m` содержит некоторое количество нулей, которое мы хотим заменить на пропущенные значения. Какая из конструкций произведёт такую замену?
  - а) `m[m == 0] <- NA`
  - б) `m <- apply(m, 1:2, function(k) if (!k) NA else k)`
  - в) `for (i in 1:nrow(m)) {for (j in 1:ncol(m)) {if (m[i, j] == 0) m[i, j] <- NA}} # omg who wrote this?`
  - г) все из вышеперечисленных
  
3. И матрица, и дата фрейм имеют прямоугольную форму, но дата фрейм может содержать данные разных типов. Так, если у нас есть матрица `mat` и дата фрейм `df`, то операция `as.data.frame(mat)` всегда будет работать корректно. А что насчёт обратной операции, `as.matrix(df)`?
  - а) такая операция запрещена в любом случае
  - б) такая операция корректно отработает только в том случае, если все переменные имеют один и тот же тип

- в) такая операция определена корректно: сработает автоматическое приведение типов
  - г) такая операция может иметь разный результат в зависимости от платформы и операционной системы
4. Какими из нижеуказанных способов можно выбрать только те строки, которые соответствуют департаментам с рейтингом (rating) ниже пятидесяти, при этом сохранив все столбцы, кроме rating?
- а) `attitude[attitude$rating < 50, -"rating"]`
  - б) `subset(sel = -rating, sub = rating < 50, attitude)`
  - в) `subset(attitude, rating < 50, -rating)`
  - г) `attitude[attitude$rating < 50, names(attitude) != "rating"]`
  - д) `attitude[rating < 50, names(attitude) != "rating"]`
5. Выберите верные утверждения:
- а) функция `strsplit` возвращает строковый вектор с результатом разбиения аргумента `x` по шаблонам, заданным аргументом `split`?
  - б) `paste0 --` это, по сути, `paste` с аргументом `sep = ""`?
  - в) если к строке сначала применить `toupper`, а потом `tolower`, то результатом будет исходная строка?
  - г) в регулярных выражениях символ знака вопроса "?" означает "один любой символ"?
6. Фактор – любопытный объект, отвечающий понятию категориальной переменной в статистике. Что вам известно о нем?
- а) по умолчанию опция `stringsAsFactors` равна `TRUE`
  - б) отношения вида "больше-меньше" определены для всех факторов
  - в) нестрого говоря, фактор – это целочисленный вектор плюс строковый вектор меток (уровней)
  - г) упорядоченный фактор можно создать функцией `order`
7. Рассмотрим "полный" вызов функции `seq` (`x`, `y` и `z` -- числа):  
`seq(from = x, to = y, by = z)`  
 Какие из следующих вызовов эквивалентны этому?
- а) `seq(b = z, f = x, t = y)`
  - б) `seq(to = y, by = z, from = x)`
  - в) `seq(by = z, x, y)`
  - г) `seq(y, z, fr = x)`
  - д) `seq(x, y, z)`
8. Какой из функций нет в базовом R?
- а) `vapply`
  - б) `eapply`
  - в) `rapply`
  - г) `mapply`

- д) tapply
- е) apply
- ж) lapply
- з) sapply
- и) dapply

9. Какие значения будут получены в ходе исполнения следующего скрипта?

```
f <- function(y) {
  y <- x + y
  y
}
```

```
g <- function(x) {
  y <- f(x)
  f <- function(x) {
    y - x
  }
  y - f(x)
}
```

```
x <- 10
y <- 1
f(x); f(y)
g(x); g(y)
x; y
```

- а) будет ошибка, потому что в локальном окружении f нет объекта с именем x
- б) 20, 11, 20, 11, 10, 1
- в) 20, 11, 10, 1, 10, 1
- г) 20, 11, 10, 1, 10, 20
- д) будет ошибка, потому что функция f объявлена раньше, чем объект с именем x

10. Выберите верные утверждения:

- а) в tidy data нет пропущенных значений
- б) колонки tidy data фрейма соответствуют наблюдениям, ряды – переменным
- в) колонки tidy data фрейма соответствуют переменным, ряды – наблюдениям
- г) для приведения к виду tidy data обязательно использовать tidyr

11. Какие из этих записей функции select пакета dplyr правильно использованы, чтобы получить только столбцы, содержащие имя, фамилию и дату?

- а) select(df, contains("name"), date\_added)



- б) `select(df, first_name, last_name, date_added)`
- в) `select(df, -3:4)`
- г) `select(df, -contains("_add"))`
- д) `df %>% select(c(1:2, 5))`
- е) `select(df, matches("_.{4,5}$"))`

12. Укажите, какая команда создаст сабсет данных `mtcars`, только для тех автомобилей, у которых число цилиндров (`cyl`) не равняется 3, и время разгона автомобиля (`qsec`) больше среднего по выборке.

- а) `new_data <- subset(mtcars, cyl == 3 | qsec > mean(qsec))`
- б) `new_data <- subset(mtcars, cyl == 3 & qsec < mean(qsec))`
- в) `new_data <- subset(mtcars, cyl != 3 | qsec < mean(qsec))`
- г) `new_data <- subset(mtcars, cyl != 3 & qsec > mean(qsec))`

13. Предположим, вы рассчитали стандартное отклонение вашей выборки, сохранили результат в переменную `my_sd` и хотите сохранить этот объект в файл `"my_sd.RData"`.

Какая команда позволит вам это сделать?

- а) `write.csv(my_sd, file = "my_sd.RData")`
- б) `rm(my_sd)`
- в) `write.table(my_sd, file = "my_sd.RData")`
- г) `save(my_sd, file = "my_sd.RData")`

14. Укажите формулы, которые расшифровываются одинаково:

- а)  $DV \sim (IV1 + IV2)^2$
- б)  $DV \sim (IV1 * IV2)^2$
- в)  $DV \sim IV1:IV2 + \text{Error}(\text{subject}/(IV1:IV2))$
- г)  $DV \sim IV1 * IV2$

15. Постройте линейную модель, в которой в качестве зависимой переменной выступает расход топлива (`mpg`), а в качестве независимых - вес машины (`wt`) и коробка передач (модифицированная `am`), а также их взаимодействие. Выведите `summary` этой модели.

Что отражает значение `intercept` в данной модели?

- а) средний расход топлива у машин с нулевым весом и ручной коробкой передач
- б) расход топлива у машин с нулевым весом
- в) расход топлива у машин со средним весом
- г) средний расход топлива у машин с автоматической коробкой передач
- д) расход топлива у машин с автоматической коробкой передач и нулевым весом

16. Постройте линейную модель, в которой в качестве зависимой переменной выступает расход топлива (`mpg`), а в качестве независимых - вес машины (`wt`)

и коробка передач (модифицированная `am`), а также их взаимодействие. Выведите `summary` этой модели.

Какие утверждения мы можем сделать на основе данной модели?

- а) в машинах с ручной коробкой передач с увеличением веса растёт расход топлива, а с автоматической - наоборот
- б) в машинах с ручной коробкой передач вес сильнее влияет на расход топлива
- в) в машинах с автоматической коробкой передач вес не влияет на расход топлива, а с ручной - влияет
- г) у машин с ручной коробкой передач расход топлива выше
- д) у машин с ручной коробкой передач расход топлива ниже
- е) вес автомобиля не влияет на расход топлива

17. В переменной `my_vector` хранится вектор значений. Какое преобразование позволяет сделать его распределение нормальным (согласно `shapiro.test`)?

- а) `log(my_vector)`
- б) все преобразования работают
- в) `sqrt(my_vector)`
- г) `1/my_vector`
- д) ни одно преобразование не работает

18. Какую из этих команд вы сможете выполнить, не загружая дополнительные пакеты?

- а) `plot(cars)`
- б) `ggplot(cars, aes(speed, dist)) + geom_point()`

19. Что произойдёт, если дать функции `ggplot` данные, но не объяснить, какой именно тип графика нужно отрисовать? Например, так:

```
ggplot(cars, aes(speed, dist))
```

- а) будет отрисован график по умолчанию (диаграмма рассеяния)
- б) будет отрисован график по умолчанию (линейная диаграмма)
- в) `r` выведет ошибку
- г) будут отрисованы подписанные оси с нанесёнными числами, но поле внутри будет пустым

20. В каком из случаев столбцы имеет смысл раскрашивать в разные цвета?

- а) для гистограммы
- б) всегда бессмысленно
- в) для столбчатой диаграммы (`bar plot`)
- г) может иметь смысл в обоих случаях

21. Что произойдёт, если не указать название файла при вызове функции `svg()` или `png()`?

- а) график не будет сохранён в файл, но будет показан в правой нижней части окна

- б) график будет сохранён в файл Rplot001 с соответствующим расширением, а при каждом следующем вызове функции этот файл будет перезаписан
- в) произойдёт ошибка, и ничего не будет сохранено
- г) график будет сохранён в файл Rplot001 с соответствующим расширением, а при каждом следующем вызове функции будут сохраняться файлы Rplot002, Rplot003 и так далее

22. Отличается ли результат работы этих двух команд?

```
ggplot(iris, aes(Species, Petal.Width, fill = Species)) + geom_boxplot()  
ggplot(iris, aes(Species, Petal.Width)) + geom_boxplot(aes(fill = Species))
```

- а) да
- б) нет

23. Вы нашли коэффициент корреляции между ростом человека в см и весом в килограммах. Пусть он составляет 0.7. Изменится этот коэффициент, если рост перевести в метры?

- а) да
- б) нет

24. Примерно 20% американцев курят. Если мы случайным образом выберем 40 американцев, какова вероятность что 10 из них курят?

- а) 40%
- б) 20%
- в) 11%
- г) 7%

25. Какая из центральных тенденций лучше подойдет для описания ряда: 2, 17, 19, 20, 18

- а) среднее значение
- б) мода
- в) медиана
- г) нет правильного ответа

26. Какой тип имеет переменная «пол»?

- а) количественная дискретная
- б) количественная непрерывная
- в) категориальная упорядоченная
- г) категориальная неупорядоченная
- д) нет правильного ответа

27. Производитель шин утверждает, что они будут служить по крайней мере 40 000 км. Тест для 40 шин показал, что среднее время службы одной шины составляет 39 750 км, среднее отклонение 387 км. Можем ли мы считать, что

среднее время пригодности шин отличается от задекларированных 40000 для  $\alpha = 0.05$ ?

- а) да
- б) нет

28. Производитель шин утверждает, что они будут служить по крайней мере 40 000 км. Тест для 40 шин показал, что среднее время службы одной шины составляет 39 750 км, среднее отклонение 387 км. Найдите тестовую статистику.

- а)  $t = -4.09$
- б)  $t = 3.23$
- в)  $t = -3.23$
- г)  $t = 2.067$
- д)  $t = 4.09$

29. Пусть  $n = 100$ , и мы хотим оценить значение среднего генеральной совокупности  $\mu = 20$ . Альтернативная гипотеза утверждает, что  $\mu \neq 20$ . Среднее значение выборки  $\bar{x} = 18$ , среднее отклонение  $s = 10$ . Вычислить p-value.

- а) 0.0228
- б) 0.0482
- в) 0.5532
- г) 0.04
- д) 0

30. При выполнении теста на значимость для пропорции, t-распределение для расчета тестовой статистики используется:

- а) при  $n$  менее 30
- б) при  $n$  больше 30
- в) всегда
- г) никогда

31. Сопоставьте следующие бинарные операторы языка с их названиями.

|     |                                      |
|-----|--------------------------------------|
| <-  | Деление                              |
| <<- | Остаток от деления                   |
| <=  | Возведение в степень                 |
| ^   | Меньше либо равно                    |
| :   | Создание числовых векторов с шагом 1 |
| /   | Оператор присваивания                |
|     | Оператор глобального присваивания    |

32. Укажите иерархию среди четырёх основных типов, выстроив их в порядке сверху вниз от самого "узкого" к самому "широкому".

1. Логический (logical)
2. Целочисленный (integer)
3. Числа с плавающей точкой (numeric/double)

#### 4. Строковый (character)

33. Сопоставьте перечисленные способы и наиболее подходящие к ним характеристики.

|               |   |
|---------------|---|
| Положительное | Способ целочисленного индексирования, для которого длина результата редко когда равна длине индекса |
| Логическое    | Способ, при котором отсутствие совпадения ведёт к созданию NA                                       |
| По именам     | Способ, при котором отсутствие совпадения не ведёт к созданию NA                                    |
| Отрицательное | Способ целочисленного индексирования, для которого длина результата всегда равна длине индекса      |
|               | Единственный способ, для которого работают правила переписывания                                    |

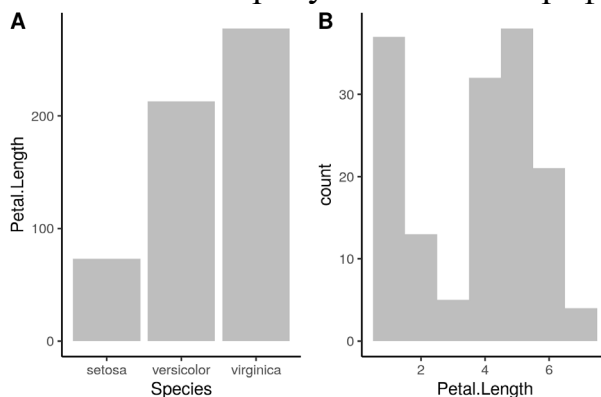
34. В каком порядке происходит разбор аргументов функции? От первого этапа к последнему.

1. Разбираются аргументы по полному названию
2. Разбираются аргументы по префиксу
3. Разбираются аргументы по позиции
4. Неразобранные аргументы передаются в ellipsis (. . .)

35. Соотнесите названия функций пакета dplyr и их описание

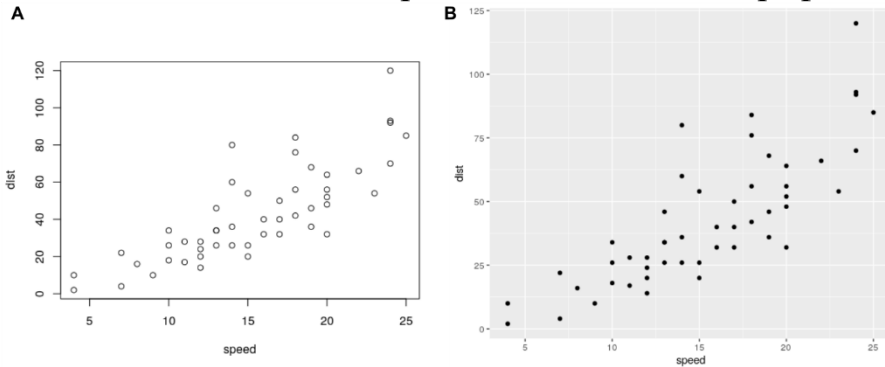
|            |   |
|------------|---|
| transmute  | То же, что и mutate, но с удалением исходных колонок  |
| sample_n   | Дата фрейм, состоящий из случайных рядов исходного  |
| inner_join | Объединение двух дата фреймов по ключу; берутся только значения по ключам, встречающимся в обоих дата фреймах |
| transform  | Какой dplyr? Это же функция из base, не надо меня путать!   |
|            | То же, что и mutate   |

36. Соотнесите рисунок с типом графика.



|   |                 |
|---|-----------------|
| A | базовая графика |
| B | ggplot2         |

37. Каким способом построен каждый из этих графиков?



|   |                      |
|---|----------------------|
| A | столбчатая диаграмма |
| B | гистограмма          |
|   | линейчатая диаграмма |

38. Справа указаны семантические особенности при работе с аргументами. Какие из вызовов слева наиболее полно отражают суть каждой из этих особенностей?

|  |   |
|--|---|
| <code>paste("number", LETTERS[1:3], 1:6, sep = " ", collapse = " and ")</code> | Использование аргументов по умолчанию                                 |
| <code>lapply(1:4, rnorm, mean = 10, sd = 10)</code>                            | Многоточие как возможность указать произвольное количество аргументов |
| <code>sort(rnorm(5))</code>  | Многоточие как "проброс аргументов"                                   |
|  | Использование правил разбора аргументов с частичным дополнением       |

39. Сопоставьте названия функций и описание того, что они делают:

|                        |   |
|------------------------|---|
| <code>replicate</code> | Применение заданной функции к всевозможным комбинациям двух аргументов  |
| <code>mapply</code>    | Запуск одного и того же выражения некоторое количество раз  |
| <code>outer</code>     | Применение заданной функции к списку аргументов   |
|                        | Применение заданной функции к произвольному количеству векторов либо списков аргументов, в итерационном порядке |

40. Установите соответствие между командами базового R и пакета dplyr.

|  |  |
|--|--|
| <code>subset(avian, PDB &gt; 0, c("Observer", "PDB"))</code> | <code>summarise(group_by(avian, Observer), x = sum(PDB &gt; 0))</code> |
|--|--|

|   |   |
|---|---|
| <code>aggregate(avian[, "PDB"], list(Observer = avian[, "Observer"]), function(x) sum(x &gt; 0))</code> | <code>filter(select(avian, Observer, PDB), PDB &gt; 0)</code> |
| <code>avian[, "PDB"] &lt;- avian[, "PDB"] &gt; 0</code>   | <code>filter (avian, PDB = PDB &gt; 0)</code>                 |
|   | <code>mutate(avian, PDB = PDB &gt; 0)</code>                  |

41. Если в R сложить два вектора, имеющих длину 4 и 6, то результат будет вектором длины?

42. Воспользуемся встроенными данными `prk`, иллюстрирующими влияние применения различных удобрений на урожайность гороха (`yield`). Нашей задачей будет выяснить, существенно ли одновременное применение азота (фактор N) и фосфата (фактор P). Примените дисперсионный анализ, где будет проверяться влияние фактора применения азота (N), влияние фактора применения фосфата (P) и их взаимодействие.

В ответе укажите `p-value` для взаимодействия факторов N и P.

43. В уже созданных переменных `number_1`, `number_2` и `number_3`, сохранены целые числа. Проверьте, действительно ли сумма первых двух чисел строго больше, чем третье число. Результат сравнения (TRUE или FALSE) сохраните в новую переменную с именем `result`.

44. В уже созданной переменной `my_numbers` сохранен вектор из 20 целых чисел. Ваша задача создать новую переменную `my_numbers_2`, в которой будет сохранен 2, 5, 7, 9, 12, 16 и 20 элемент вектора `my_numbers`.

45. Создайте новую числовую переменную `new_var` в данных `mtcars`, которая содержит единицы в строчках, если в машине не меньше четырёх карбюраторов (переменная `"carb"`) или больше шести цилиндров (переменная `"cyl"`). В строчках, в которых условие не выполняется, должны стоять нули.

46. В уже существующей переменной `my_vector` сохранен вектор из 50 чисел. Решите задачу используя конструкцию:

```
if () {
} else {
}

```

Если среднее значение вектора `my_vector` больше 20, в переменную `result` сохраните "My mean is great", если среднее значение `my_vector` меньше или равно 20 то в переменную `result` сохраните строку "My mean is not so great".

47. Рассчитайте среднее значение времени разгона (`qsec`) для автомобилей, число цилиндров (`cyl`) у которых не равняется 3 и показатель количества миль на галлон топлива (`mpg`) больше 20.

Получившийся результат (среднее значение) сохраните в переменную `result`.

48. Нужно построить scatterplot с помощью ggplot из ggplot2, по оси x которого будет mpg, по оси y - disp, а цветом отобразить переменную (hp).

Полученный график нужно сохранить в переменную plot1. Таким образом в ответе должен быть скрипт:

```
plot1 <- ggplot(data, aes())+  
  geom_****()
```

49. Воспользуемся данными diamonds из библиотеки ggplot2. При помощи критерия Хи - квадрат проверьте гипотезу о взаимосвязи качества огранки бриллианта (cut) и его цвета (color). В переменную main\_stat сохраните значение статистики критерия Хи - квадрат. Обратите внимание, main\_stat должен быть вектором из одного элемента, а не списком (листом).

50. При помощи точного критерия Фишера проверьте гипотезу о взаимосвязи типа коробки передач (am) и типа двигателя (vs) в данных mtcars. Результат выполнения критерия сохраните в переменную. Получившийся p - уровень значимости сохраните в переменную fisher\_test.

51. Воспользуемся встроенным набором данных в R – ToothGrowth. Данные позволяют исследовать рост зубов у морских свинок в зависимости от дозировки витамина С и типа потребляемых продуктов.

Сравните среднее значение длины зубов свинок, которые потребляли апельсиновый сок (OJ) с дозировкой 0.5 миллиграмм, со средним значением длины зубов свинок, которые потребляли аскорбиновую кислоту (VC) с дозировкой 2 миллиграмма.

Значение t-критерия сохраните в переменную t\_stat.

52. Воспользуемся встроенными данными prk, иллюстрирующими влияние применения различных удобрений на урожайность гороха (yield). Нашей задачей будет выяснить, существенно ли одновременное применение азота (фактор N) и фосфата (фактор P). Примените дисперсионный анализ, где будет проверяться влияние фактора применения азота (N), влияние фактора применения фосфата (P) и их взаимодействие.

В ответе укажите p-value для взаимодействия факторов N и P.

53. Напишите функцию, которая выводит номера позиций пропущенных наблюдений в векторе.

На вход функция получает числовой вектор с пропущенными значениями. Функция возвращает новый вектор с номерами позиций пропущенных значений.

54. Напишите функцию corr.calc, которая на вход получает data.frame с двумя количественными переменными, рассчитывает коэффициент корреляции Пирсона и возвращает вектор из двух значений: коэффициент корреляции и p - уровень значимости.



55. Воспользуемся уже знакомыми данными diamonds из библиотеки ggplot2. Только для бриллиантов класса Ideal (переменная cut) с числом карат равным 0.46 (переменная carat) постройте линейную регрессию, где в качестве зависимой переменной выступает price, в качестве предиктора - переменная depth. Сохраните коэффициенты регрессии в переменную fit\_coef.

56. Напишите функцию fill\_na, которая принимает на вход данные с тремя переменными:

x\_1 – числовой вектор

x\_2 – числовой вектор

y – числовой вектор с пропущенными значениями.

Теперь – самое интересное. На первом этапе, используя только наблюдения, в которых нет пропущенных значений, мы построим регрессионную модель (без взаимодействий), где y – зависимая переменная, x\_1 и x\_2 – независимые переменные. Затем, используя построенную модель, мы заполним пропущенные значения предсказаниями модели.

Функция должна возвращать dataframe с новой переменной y\_full. Сохраните в нее переменную y, в которой пропущенные значения заполнены предсказанными значениями построенной модели.

57. С помощью функции step найдите оптимальную модель для предсказания rating в датасете attitude. Model\_full и model\_null уже созданы. Сохраните команду с функцией step в переменную ideal\_model.

58. Напишите функцию, которая на вход получает dataframe с двумя количественными переменными, а возвращает стандартизованные коэффициенты для регрессионной модели, в которой первая переменная датафрейма выступает в качестве зависимой, а вторая в качестве независимой.

59. Напишите функцию resid.norm, которая тестирует распределение остатков от модели на нормальность при помощи функции shapiro.test и создает гистограмму при помощи функции ggplot() с красной заливкой "red", если распределение остатков значительно отличается от нормального ( $p < 0.05$ ), и с зелёной заливкой "green" - если распределение остатков значительно не отличается от нормального.

На вход функция получает регрессионную модель. Функция возвращает переменную, в которой сохранен график ggplot.

60. Используем данные mtcars. Сохраните в переменную логистическую регрессионную модель, где в качестве зависимой переменной выступает тип коробки передач (am), в качестве предикторов переменные disp, vs, mpg. Значения коэффициентов регрессии сохраните в переменную log\_coef.

61. В переменной `my_df` сохранен `dataframe` с произвольным числом количественных переменных. При помощи функции `apply` найдите максимальное значение в каждой строке. Сохраните результат (вектор максимальных значений) в переменную `row_max`.

62. Напишите функцию `positive_sum`, которая получает на вход `dataframe` с произвольным количеством числовых переменных. Основная задача функции - найти сумму положительных значений в каждой переменной и сохранить их в список.

63. Поработаем с данными `diamonds` из пакета `ggplot2`.  
В переменную `d` сохраните только нечетные строки исходных данных `diamonds`.

64. Напишите функцию, `all_to_factor`, которая преобразует `dataframe`, переводя все его переменные в фактор.

65. Напишите функцию, `to_factors`, которая получает на вход `dataframe` с произвольным числом количественных переменных и вектор с номерами колонок, которые нужно перевести в фактор.  
Для перевода числовых колонок в фактор будем использовать следующий принцип, если наблюдение больше среднего всей переменной, то 1, иначе 0.

66. Используя функцию `qplot`, постройте гистограмму переменной `depth` из данных `diamonds`. Сохраните график в переменную `depth_hist`.

67. Постройте диаграмму рассеивания (`scatter plot`) как в указанном ниже примере, результат сохраните в переменную `price_carat_clarity_points`.  
данные – `diamonds`  
ось x – `carat`  
ось y – `price`  
цвет точек – `clarity`

68. Отобразите взаимосвязь между доходом (`income`) и числом продаж (`sale`), цветом точек указав номер магазина (`shop`):  
Сохраните график в переменную `my_plot`. Обратите внимание, что линия тренда одна для всех наблюдений. Данные хранятся в переменной `sales`.

69. Используя данные `iris`, постройте график плотности для переменной `Sepal.Length`. Разбейте график на части по переменной `Species` при помощи `facet_wrap`. Результат сохраните в переменную `sl_wrap`.

70. Возьмите данные `diamonds`. Создайте новый `dataframe` с именем `high_price`, в котором будут храниться только 10 самых дорогих бриллиантов каждого

цвета. Также в итоговом dataframe должны храниться только две переменные color и price.

#### Критерии и шкала оценивания по оценочному средству «тесты»

| Шкала оценивания<br>(интервал баллов) | Критерий оценивания   |
|---------------------------------------|---|
| 5                                     | Тесты выполнены на высоком уровне (правильные ответы даны на 90-100% тестов)                    |
| 4                                     | Тесты выполнены на среднем уровне (правильные ответы даны на 75-89% тестов)                     |
| 3                                     | Тесты выполнены на низком уровне (правильные ответы даны на 50-74% тестов)                      |
| 2                                     | Тесты выполнены на неудовлетворительном уровне (правильные ответы даны менее чем на 50% тестов) |

#### Вопросы для обсуждения (в виде докладов и сообщений)

1. Обработка информации.
2. История языка R.
3. Сферы применения, характерные особенности R.
4. Команды R.
5. Сохранение данных и удаление объектов.
6. Оператор присваивания «`()`».
7. Числа и векторы.
8. Векторная арифметика.
9. Логические векторы. Векторы символов. Векторы индексов.
10. Объекты и их режимы.
11. Массивы и матрицы.
12. Матричные инструменты.
13. Списки и фреймы данных.
14. Графические процедуры. Функция `plot()`.
15. Графические элементы. Оси и метки. Поля рисунка.
16. Выделение тренда.
17. Интерактивная графика.
18. Статистические модели в R.
19. Дисперсионный анализ и сравнение моделей. Таблицы ANOVA.
20. Нелинейный анализ временных рядов.

## Критерии и шкала оценивания по оценочному средству «вопросы для обсуждения»

| Шкала оценивания<br>(интервал баллов) | Критерий оценивания   |
|---------------------------------------|---|
| 5                                     | Доклад (сообщение) представлен(о) на высоком уровне (студент в полном объеме осветил рассматриваемую проблематику, привел аргументы в пользу своих суждений, владеет профильным понятийным (категориальным) аппаратом и т.п.) |
| 4                                     | Доклад (сообщение) представлен(о) на среднем уровне (студент в целом осветил рассматриваемую проблематику, привел аргументы в пользу своих суждений, допустив некоторые неточности и т.п.)                                    |
| 3                                     | Доклад (сообщение) представлен(о) на низком уровне (студент допустил существенные неточности, изложил материал с ошибками, не владеет в достаточной степени профильным категориальным аппаратом и т.п.)                       |
| 2                                     | Доклад (сообщение) представлен(о) на неудовлетворительном уровне или не представлен (студент не готов, не выполнил задание и т.п.)  |

### Оценочные средства для промежуточной аттестации (зачет) Теоретические вопросы

1. Объекты языка R, организация обращения к элементам данных.
2. Операторы [, [[, \$. Извлечение элементов с учетом свойств.
3. Создание скриптов. Запуск скриптов.
4. Факторы как объекты для хранения и обработки символьной информации.
5. Формулы отклик ~ воздействие, их синтаксис.
6. Особенности типов данных (NA, NaN, Inf).
7. Команды, связанные с отдельными типами распределений: r\*(), f\*(), d\*(), q\*() (здесь \* -- идентификатор конкретного распределения).
8. Создание объектов R.
9. Ввод данных с внешних носителей.
10. Вывод данных в файлы (команды sink(), print(), cat() и другие. Их особенности).
11. Графические команды высокого уровня plot(), hist(), boxplot().
12. Графические команды низшего уровня lines(), points(), text(), legend() и другие.
13. Настройка параметров вывода графической информации.
14. Команды проверки гипотез. Их параметры.
15. Условный оператор if() {}.
16. Операторы цикла for() {}, while() {}.
17. Подключение внешних библиотек. Примеры библиотек.
18. Работа со справочными материалами. Команды ?, help(), apropos()
19. Команды сортировки sort(), order().
20. Команда lm(), объект типа lm.
21. Команды кластеризации (k-means(), hclust(), cutree(...)).

Критерии и шкала оценивания по оценочному средству промежуточный контроль («зачет»)

| Характеристика знания предмета и ответов  | Зачеты     |
|---|------------|
| Студент глубоко и в полном объёме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач. | зачтено    |
| Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.                                      |            |
| Студент знает только основной программный материал, допускает неточности, недостаточно чёткие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.                                    |            |
| Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.                           | не зачтено |

### Лист изменений и дополнений

| №<br>п/п | Виды дополнений и<br>изменений | Дата и номер протокола<br>заседания кафедры<br>(кафедр), на котором<br>были рассмотрены и<br>одобрены изменения и<br>дополнения | Подпись (с<br>расшифровкой)<br>заведующего кафедрой<br>(заведующих кафедрами) |
|----------|--------------------------------|---|---|
|          |                                |   |   |
|          |                                |   |   |
|          |                                |   |   |
|          |                                |   |   |