

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания
по выполнению лабораторных работ
по дисциплине
«Программирование на R»
для студентов направления подготовки
02.03.01 Математика и компьютерные науки
направленность (профиль)
«Анализ данных и искусственный интеллект»

Лабораторная работа 1. Среда разработки для языка R: RStudio, RStudio.cloud

Теоретическая часть:

1. Ознакомление с языком R, его особенностями и возможностями

Язык R является мощным инструментом для статистического анализа данных и визуализации. Он широко используется в науке, экономике, биоинформатике и других сферах, требующих обработки больших объемов данных. R был разработан как свободная альтернатива языку S и обладает следующими особенностями:

- Открытый исходный код: язык распространяется бесплатно и поддерживается сообществом разработчиков.
- Богатая экосистема: включает в себя тысячи пакетов для анализа данных, машинного обучения и построения графиков.
- Гибкость и расширяемость: пользователи могут создавать собственные функции, подключать сторонние библиотеки и автоматизировать обработку данных.
- Визуализация данных: R предлагает широкий набор инструментов для построения графиков и диаграмм, включая библиотеку ggplot2.
- Интерактивная работа: можно работать как в командной строке, так и в интегрированной среде разработки RStudio.

Таким образом, R является универсальным языком для обработки и анализа данных, который подходит как для начинающих, так и для опытных аналитиков.

2. Среда разработки RStudio: интерфейс, установка, настройка

RStudio – это интегрированная среда разработки (IDE) для работы с языком R. Она предоставляет удобный интерфейс, который облегчает написание кода, выполнение команд и визуализацию данных.

Интерфейс RStudio

RStudio состоит из нескольких основных панелей:

- Source Editor – область для написания и редактирования кода.
- Console – командная строка для непосредственного выполнения команд.
- Environment/History – вкладка для управления переменными и просмотра истории команд.
- Files/Plots/Packages/Help – секция для работы с файлами, графиками, пакетами и документацией.

Установка R и RStudio

1. Загрузка и установка R: необходимо скачать последнюю версию R с официального сайта [CRAN](https://cran.r-project.org/).
2. Загрузка и установка RStudio: IDE RStudio можно загрузить с [официального сайта RStudio](https://www.rstudio.com/).
3. Настройка RStudio: после установки рекомендуется настроить рабочую директорию, включить перенос строк и выбрать подходящую цветовую схему.

Основные настройки и работа

- Создание нового R-скрипта через File > New File > R Script.
- Выполнение кода с помощью комбинации клавиш Ctrl + Enter.
- Установка дополнительных пакетов через Tools > Install Packages или с использованием `install.packages("название_пакета")` в консоли.
- Управление рабочей средой и переменными через панель Environment.

Использование RStudio делает процесс работы с R удобнее, обеспечивая мощные инструменты для анализа и визуализации данных.

3. Работа в RStudio.cloud

RStudio.cloud – это облачная версия RStudio, позволяющая работать с R через веб-браузер без необходимости установки программного обеспечения на локальный компьютер. Это особенно полезно для учебных целей, так как предоставляет единообразную среду для всех пользователей.

Преимущества RStudio.cloud:

- Доступность – не требует установки, работает через браузер.
- Облачное хранение – проекты и данные сохраняются в облаке и доступны с любого устройства.
- Легкость управления зависимостями – автоматическая настройка необходимой среды для работы с R.
- Совместная работа – возможность совместного редактирования проектов.

Работа с RStudio.cloud:

1. Регистрация и вход в систему: Перейдите на сайт [RStudio.cloud](https://rstudio.cloud/) и создайте учетную запись.
2. Создание нового проекта: Нажмите New Project, выберите R Script для создания нового R-скрипта.
3. Работа с кодом: Пишите и выполняйте команды в консоли или в скриптовом файле.

4. Установка пакетов: Используйте `install.packages("название_пакета")` для установки дополнительных пакетов.
5. Сохранение и экспорт результатов: Вы можете загружать и скачивать файлы через панель Files.

Использование RStudio.cloud делает обучение и работу с R удобными и доступными без привязки к конкретному рабочему месту.

Практическая часть

1. Установить R и RStudio (или зарегистрироваться в RStudio.cloud).
2. Создать новый проект и выполнить базовые команды.
3. Изучить интерфейс RStudio.

Контрольные вопросы

1. Какие основные панели присутствуют в интерфейсе RStudio?
 2. Как создать и выполнить R-скрипт?
 3. Где можно изменить настройки рабочей среды?
 4. Как загрузить пакеты в RStudio?
 5. Как сохранить и загрузить рабочую сессию?
-

Лабораторная работа 2. Типы данных и структуры в R

Теоретическая часть:

1. Основные типы данных в R

Тип данных определяет, какие значения может хранить переменная и какие операции над ней допустимы. В R различают следующие основные типы данных:

1.1 Числовые данные (Numeric)

- Представляют собой числа с плавающей запятой (double) или целые числа (integer).
- Используются в математических операциях: сложение, вычитание, умножение, деление.
- Пример: `x <- 10.5`, `y <- 5L` (где L указывает на целое число).

1.2 Строковые данные (Character)

- Используются для хранения текста.
- Можно объединять, разделять, изменять регистр.
- Пример: `name <- "Привет, R!"`.

1.3 Логические данные (Logical)

- Принимают два значения: TRUE и FALSE.
- Используются в логических выражениях и фильтрации данных.
- Пример: `is_active <- TRUE`.

1.4 Факторы (Factor)

- Представляют категориальные переменные.
- Имеют уровни (levels), которые можно упорядочить.
- Пример: `factor_var <- factor(c("Мужской", "Женский", "Мужской"))`.

1.5 Комплексные числа (Complex)

- Содержат мнимую единицу i.
- Используются в математических вычислениях.
- Пример: `z <- 2 + 3i`.

2. Структуры данных в R

R предлагает несколько структур данных для организации информации.

2.1 Векторы (Vectors)

- Одномерные массивы элементов одного типа.
- Пример: `v <- c(1, 2, 3, 4)`.

2.2 Матрицы (Matrices)

- Двумерные массивы элементов одного типа.
- Пример: `m <- matrix(1:9, nrow=3, ncol=3)`.

2.3 Списки (Lists)

- Структура, содержащая элементы разного типа.
- Пример: `lst <- list(число=10, текст="Привет", логика=TRUE)`.

2.4 Таблицы данных (Data Frames)

- Табличный формат хранения данных.
- Используется в анализе данных.
- Пример: `df <- data.frame(Имя=c("Анна", "Иван"), Возраст=c(25, 30))..`

3. Работа с факторами

Факторы в R представляют собой категориальные переменные, которые используются для хранения данных, принимающих ограниченное количество значений (уровней). Они полезны при анализе данных, особенно в статистическом моделировании, так как позволяют кодировать категориальные переменные оптимизированным способом.

Особенности факторов

- Каждый фактор имеет фиксированный набор уровней (levels).
- Уровни могут быть упорядоченными (например, "низкий", "средний", "высокий") или неупорядоченными (например, "мужской", "женский").
- Факторы используются в статистических моделях и позволяют корректно обрабатывать категориальные переменные.

Создание факторов

Факторы создаются с помощью функции `factor()`, например:

```
f <- factor(c("Красный", "Синий", "Зеленый", "Красный", "Синий"))  
print(f)  
levels(f)
```

Изменение уровней факторов

Можно задать уровни вручную с помощью параметра levels:

```
f <- factor(c("Низкий", "Средний", "Высокий", "Средний"),  
           levels = c("Низкий", "Средний", "Высокий"), ordered = TRUE)
```

Этот фактор будет упорядоченным, что позволит проводить сравнения (<, >).

Преобразование факторов в другие типы

Фактор можно преобразовать в строку или числовой тип:

```
as.character(f)  
as.numeric(f)
```

Факторы играют важную роль в обработке данных и широко используются в статистическом анализе.

Практическая часть

1. Создать объекты разных типов данных.
2. Преобразовать один тип данных в другой.
3. Создать векторы, матрицы и списки.

Контрольные вопросы

1. Какие основные типы данных поддерживаются в R?
 2. В чем отличие векторов и списков?
 3. Как изменить тип данных переменной?
 4. Как создать фактор и задать уровни?
 5. Как узнать размерность матрицы?
-

Лабораторная работа 3. Импорт данных и построение срезов

Теоретическая часть

1. Способы загрузки данных в R

В R предусмотрены несколько способов загрузки данных из различных источников, таких как текстовые файлы, электронные таблицы, базы данных и API. В зависимости от формата данных используются разные функции.

1.1 Чтение данных из CSV-файлов

CSV (Comma-Separated Values) – один из самых распространенных форматов хранения табличных данных. Для его загрузки используется функция `read.csv()`:

```
my_data <- read.csv("data.csv", header = TRUE, sep = ",")
```

Где:

- `header = TRUE` – указывает, что первая строка содержит названия столбцов.
- `sep = ","` – задает разделитель значений.

1.2 Чтение данных из Excel

Для работы с файлами Excel используются пакеты `readxl` и `openxlsx`.

```
library(readxl)  
data <- read_excel("data.xlsx", sheet = 1)
```

Здесь `sheet = 1` указывает номер листа в файле.

1.3 Чтение данных из JSON

Для работы с JSON используется пакет `jsonlite`.

```
library(jsonlite)  
data <- fromJSON("data.json")
```

1.4 Импорт данных из базы данных

Подключение к базам данных выполняется через пакеты `DBI` и `RMySQL`.

```
library(DBI)  
conn <- dbConnect(RMySQL::MySQL(), dbname = "database_name", host =  
"localhost", user = "user", password = "password")
```

```
data <- dbGetQuery(conn, "SELECT * FROM table_name")
dbDisconnect(conn)
```

Эти методы позволяют загружать данные в R из различных источников для дальнейшего анализа.

2. Форматы данных: CSV, Excel, JSON

В R поддерживаются различные форматы данных, которые можно импортировать и экспортировать для анализа. Основные форматы включают:

2.1 CSV (Comma-Separated Values)

CSV – один из самых популярных форматов хранения данных. Каждая строка представляет собой запись, а значения разделяются запятыми или другими символами (такими как ; или).

- Преимущества:
 - Простота использования
 - Совместимость с различными программами (Excel, базы данных, текстовые редакторы)
- Чтение CSV-файлов в R выполняется с помощью `read.csv()`:

```
my_data <- read.csv("data.csv", header = TRUE, sep = ",")
```

- Запись в CSV осуществляется с помощью `write.csv()`:

```
write.csv(my_data, "output.csv", row.names = FALSE)
```

2.2 Excel (XLS/XLSX)

Файлы Excel более удобны для представления данных благодаря поддержке форматирования и формул. Для работы с ними используются пакеты `readxl` и `openxlsx`:

- Чтение данных из Excel-файла:

```
library(readxl)
data <- read_excel("data.xlsx", sheet = 1)
```

- Запись данных в Excel-файл:

```
library(openxlsx)
write.xlsx(data, "output.xlsx")
```

2.3 JSON (JavaScript Object Notation)

JSON – это формат хранения структурированных данных, широко используемый для передачи данных между веб-приложениями и серверами.

- Чтение JSON в R выполняется с помощью пакета jsonlite:

```
library(jsonlite)
data <- fromJSON("data.json")
```

- Запись данных в JSON-файл:

```
write_json(data, "output.json")
```

Каждый из этих форматов имеет свои преимущества и применяется в зависимости от целей анализа данных.

3. Фильтрация и подвыборки

Фильтрация данных – это процесс выбора определенных строк или столбцов из набора данных на основе заданных условий. В R этот процесс выполняется с помощью логических выражений и функций для работы с таблицами данных.

3.1 Фильтрация строк по условиям

Для фильтрации строк в data.frame используется логическая индексация.

```
# Выбор строк, где значение столбца Age больше 30
data[data$Age > 30, ]
```

Фильтрацию можно комбинировать с операторами & (и) и | (или):

```
# Выбор строк, где Age больше 30 и Salary превышает 50000
data[data$Age > 30 & data$Salary > 50000, ]
```

3.2 Выбор столбцов

Для выбора конкретных столбцов используется индексация по имени или номеру:

```
# Выбор столбцов Name и Salary
data[, c("Name", "Salary")]
```

3.3 Использование функции subset()

Функция `subset()` позволяет фильтровать строки и выбирать столбцы более читаемым способом.

```
# Выбор строк с Age > 30 и только Name и Salary  
subset(data, Age > 30, select = c(Name, Salary))
```

3.4 Использование `dplyr` для фильтрации

Пакет `dplyr` предлагает удобный синтаксис для работы с данными.

```
library(dplyr)  
data_filtered <- data %>% filter(Age > 30) %>% select(Name, Salary)
```

Фильтрация и подвыборки помогают работать с нужными данными и исключать ненужные записи для последующего анализа.

Практическая часть

1. Импортировать данные из CSV-файла.
2. Сделать подвыборку данных по критериям.
3. Экспортировать данные в файл.

Контрольные вопросы

1. Какие функции используются для загрузки CSV-файлов?
 2. Как выбрать определенные строки или столбцы из таблицы?
 3. Как фильтровать данные по условиям?
 4. Как сохранить обработанные данные в файл?
 5. Чем отличается `read.csv` от `read.table`?
-

Лабораторная работа 4. Описательный анализ данных

Теоретическая часть

1. Основные статистические параметры

Описательный анализ данных включает в себя расчет основных статистических показателей, которые позволяют определить ключевые характеристики распределения данных.

1.1 Среднее значение (Mean)

- Среднее арифметическое вычисляется как сумма всех значений, деленная на их количество.
- В R используется функция `mean()`:

```
mean(data$column_name)
```

1.2 Медиана (Median)

- Центральное значение отсортированного ряда данных.
- В R вычисляется с помощью `median()`:

```
median(data$column_name)
```

1.3 Дисперсия (Variance) и стандартное отклонение (Standard Deviation)

- Дисперсия показывает степень разброса данных относительно среднего значения.
- Стандартное отклонение является квадратным корнем из дисперсии.
- В R вычисляется с помощью `var()` и `sd()`:

```
var(data$column_name)
```

```
sd(data$column_name)
```

1.4 Квартили и межквартильный размах

- Квартили делят данные на четыре части:
 - Q1 (первый квартиль) – 25%
 - Q2 (медиана) – 50%
 - Q3 (третий квартиль) – 75%
- Межквартильный размах (IQR) – разница между третьим и первым квартилями.
- В R вычисляется с помощью `quantile()` и `IQR()`:

```
quantile(data$column_name)
```

```
IQR(data$column_name)
```

Использование этих показателей помогает получить первичную информацию о наборе данных и выявить возможные выбросы.

2. Функции для описательного анализа в R

R предоставляет широкий набор функций для описательного анализа данных, позволяя быстро получать основные статистические характеристики.

2.1 Основные функции для расчета статистических параметров

- `mean(x)`: вычисляет среднее значение вектора `x`.
- `median(x)`: вычисляет медиану вектора `x`.
- `sd(x)`: вычисляет стандартное отклонение.
- `var(x)`: вычисляет дисперсию данных.
- `summary(x)`: предоставляет краткое описание данных, включая среднее, медиану, минимум, максимум и квартили.
- `quantile(x, probs)`: вычисляет определенные квантили.

Пример использования:

```
x <- c(5, 8, 12, 15, 20)
mean(x) # Среднее значение
sd(x)   # Стандартное отклонение
summary(x) # Общий описательный анализ
```

2.2 Работа с пропущенными значениями

- `is.na(x)`: проверяет наличие пропущенных значений.
- `na.omit(x)`: удаляет строки с пропущенными значениями.
- `complete.cases(x)`: выбирает только полные наблюдения.

Пример обработки пропущенных значений:

```
x <- c(10, NA, 20, 30, NA, 40)
x_clean <- na.omit(x) # Удаление NA-значений
```

2.3 Визуализация описательной статистики

- `boxplot(x)`: строит коробчатую диаграмму (`boxplot`) для визуального анализа данных.
- `hist(x)`: строит гистограмму распределения данных.

Пример построения гистограммы:

```
hist(x, main="Гистограмма распределения", xlab="Значения", col="blue")
```

Эти функции позволяют быстро анализировать данные, выявлять выбросы и понимать их распределение.

Практическая часть

1. Рассчитать основные статистики для данных.
2. Найти пропуски и обработать их.
3. Вывести сводную информацию по набору данных.

Контрольные вопросы

1. Какие функции используются для расчета среднего и медианы?
 2. Как найти минимум и максимум в данных?
 3. Какие функции позволяют визуализировать распределение данных?
 4. Как обработать пропущенные значения?
 5. Как вывести сводную статистику для набора данных?
-

Лабораторная работа 5. Разведочный анализ данных

Теоретическая часть

1. Методы анализа распределения данных

Анализ распределения данных позволяет понять, как значения переменной распределены в выборке, выявить закономерности, выбросы и определить тип распределения.

1.1 Описательные статистики

- **Среднее (mean)** – показывает центральное значение выборки.
- **Медиана (median)** – разделяет данные на две равные части.
- **Мода (mode)** – наиболее часто встречающееся значение.
- **Дисперсия (variance) и стандартное отклонение (standard deviation)** – характеризуют степень разброса данных относительно среднего.
- **Квартили (quartiles) и межквартильный размах (IQR)** – позволяют определить разброс значений и выявить выбросы.

Пример вычисления в R:

```
summary(data$column)
var(data$column)
sd(data$column)
quantile(data$column, probs = c(0.25, 0.5, 0.75))
```

1.2 Графические методы анализа распределения

- **Гистограмма (histogram)** – показывает частоту встречаемости значений.
- **Ядерная оценка плотности распределения (density plot)** – позволяет оценить сглаженное распределение данных.
- **Коробчатая диаграмма (boxplot)** – помогает выявить выбросы и распределение значений.

Пример построения гистограммы и boxplot в R:

```
hist(data$column, main="Гистограмма", xlab="Значения", col="lightblue",
breaks=20)
boxplot(data$column, main="Boxplot", horizontal=TRUE)
```

1.3 Проверка нормальности распределения

- **Графический метод** – сравнение гистограммы с нормальным распределением.

- **Тест Шапиро-Уилка (Shapiro-Wilk test)** – используется для проверки гипотезы о нормальности данных:

```
shapiro.test(data$column)
```

- **Квантиль-квантиль график (Q-Q plot)** – позволяет визуально оценить отклонение от нормального распределения:

```
qqnorm(data$column)
qqline(data$column, col = "red")
```

1.4 Выявление выбросов

Выбросы – это значения, которые значительно отличаются от остальных наблюдений в выборке. Методы выявления:

- Использование межквартильного размаха (IQR):

```
Q1 <- quantile(data$column, 0.25)
Q3 <- quantile(data$column, 0.75)
IQR <- Q3 - Q1
outliers <- data$column[data$column < (Q1 - 1.5 * IQR) | data$column > (Q3 + 1.5 * IQR)]
```

- Построение boxplot для наглядной идентификации выбросов.

Использование этих методов позволяет глубже понять структуру данных и их распределение, что является важным этапом предварительного анализа.

2. Гистограммы и коробчатые диаграммы

Гистограммы и коробчатые диаграммы являются важными инструментами анализа распределения данных. Они помогают визуализировать структуру данных, выявить выбросы и определить тип распределения.

2.1 Гистограммы (Histogram)

Гистограмма показывает частоту значений переменной в виде столбцов. Это один из самых простых способов понять распределение данных.

- **Создание гистограммы в R:**

```
hist(data$column, main="Гистограмма", xlab="Значения", col="lightblue", breaks=20)
```

- **Параметры функции hist():**
 - o main – заголовок графика.

- o xlab – подпись оси X.
- o col – цвет столбцов.
- o breaks – количество интервалов (столбцов).

2.2 Коробчатые диаграммы (Boxplot)

Коробчатая диаграмма (или диаграмма размаха) – это метод отображения распределения данных с помощью медианы, квартилей и выбросов.

- **Создание boxplot в R:**

```
boxplot(data$column, main="Boxplot", horizontal=TRUE, col="orange")
```

- **Параметры функции boxplot():**

- o main – заголовок.
- o horizontal – горизонтальное расположение.
- o col – цвет диаграммы.

2.3 Сравнение гистограмм и boxplot

- Гистограмма показывает распределение частоты значений, но не указывает выбросы.
- Коробчатая диаграмма позволяет определить выбросы и симметрию распределения.

Использование этих методов вместе позволяет получить полную картину о распределении данных.

3. Выбросы в данных

Выбросы (outliers) – это значения, которые значительно отличаются от остальных данных в выборке. Они могут быть результатом ошибок ввода, шумов в данных или реальных аномалий, требующих анализа. Определение и обработка выбросов важны для корректного статистического анализа.

3.1 Методы выявления выбросов

1. **Метод межквартильного размаха (IQR)**

- o Вычисление первого (Q1) и третьего (Q3) квартилей.
- o Расчет межквартильного размаха ($IQR = Q3 - Q1$).
- o Определение выбросов как значений, выходящих за границы:
 - Нижняя граница: $Q1 - 1.5 * IQR$
 - Верхняя граница: $Q3 + 1.5 * IQR$
- o Пример в R:

2. `Q1 <- quantile(data$column, 0.25)`

3. `Q3 <- quantile(data$column, 0.75)`

4. $IQR <- Q3 - Q1$
5. $outliers <- data\$column[data\$column < (Q1 - 1.5 * IQR) | data\$column > (Q3 + 1.5 * IQR)]$
6. **Графический анализ (boxplot)**
 - o Коробчатая диаграмма помогает визуально определить выбросы.
 - o Пример в R:
7. $boxplot(data\$column, main="Выбросы в данных")$
8. **Z-оценка (стандартное отклонение)**
 - o Выбросы определяются как значения, отклоняющиеся от среднего более чем на 3 стандартных отклонения.
 - o Пример в R:
9. $z_scores <- scale(data\$column)$
10. $outliers <- data\$column[abs(z_scores) > 3]$

3.2 Обработка выбросов

- **Удаление выбросов** – подходит, если выбросы являются ошибками измерения.
- **Замена выбросов на медиану** – помогает сохранить структуру данных.
- **Применение логарифмического преобразования** – уменьшает влияние выбросов на анализ.

Пример замены выбросов на медиану в R:

```
median_value <- median(data$column, na.rm = TRUE)
data$column[data$column < (Q1 - 1.5 * IQR) | data$column > (Q3 + 1.5 * IQR)]
<- median_value
```

Выбросы могут существенно влиять на результаты анализа, поэтому важно правильно их идентифицировать и обрабатывать в зависимости от контекста исследования.

Практическая часть

1. Построить гистограмму и boxplot для данных.
2. Найти выбросы и обработать их.
3. Оценить нормальность распределения.

Контрольные вопросы

1. Какие графики используются для анализа распределения?
2. Как определить выбросы в данных?
3. Какие функции помогают анализировать нормальность?
4. Как вычислить основные моменты распределения?
5. Какие методы устранения выбросов существуют?

Лабораторная работа 6. Проверка статистических гипотез

Теоретическая часть

1. Основы статистических тестов

Статистические тесты – это методы, используемые для проверки гипотез о данных. Они помогают определить, значимы ли различия между выборками, соответствуют ли данные определённому распределению, или есть ли взаимосвязь между переменными.

1.1 Основные понятия статистического тестирования

- **Нулевая гипотеза (H_0)** – утверждение, предполагающее отсутствие эффекта или различий.
- **Альтернативная гипотеза (H_1)** – утверждение, противоположное нулевой гипотезе.
- **Уровень значимости (α)** – вероятность ошибочного отклонения H_0 (обычно 0.05).
- **p-value** – вероятность получения наблюдаемых данных при условии, что H_0 верна.
- **Статистическая мощность теста** – вероятность корректного отклонения H_0 , если H_1 истинна.

1.2 Виды статистических тестов

Статистические тесты делятся на параметрические и непараметрические:

1. **Параметрические тесты** – применяются, когда данные соответствуют определённым распределениям (например, нормальному).
 - t-критерий Стьюдента (t-test)
 - ANOVA (дисперсионный анализ)
 - Корреляционный анализ (Pearson)
2. **Непараметрические тесты** – используются, когда данные не соответствуют предположениям параметрических тестов.
 - U-критерий Манна-Уитни
 - Критерий Вилкоксона
 - Критерий хи-квадрат

1.3 Проверка нормальности распределения

Перед применением параметрических тестов необходимо проверить, соответствует ли распределение данных нормальному:

- **Графический метод** – гистограмма, QQ-plot.
- **Статистические тесты:**

```
shapiro.test(data$column) # Тест Шапиро-Уилка
ks.test(data$column, "pnorm") # Критерий Колмогорова-Смирнова
```

Если данные не являются нормально распределёнными, рекомендуется использовать непараметрические методы.

Использование статистических тестов позволяет делать обоснованные выводы о данных и их взаимосвязях.

2. Тесты для одной и двух выборок

Статистические тесты применяются для сравнения выборок и проверки различий между ними. Они делятся на тесты для одной выборки и тесты для двух независимых или зависимых выборок.

2.1 Тесты для одной выборки

Используются для проверки того, отличается ли среднее значение выборки от заданного значения или ожидаемого распределения.

- **t-тест для одной выборки (One Sample t-test)** – сравнивает среднее значение выборки с теоретическим.
- **Тест знаков (Sign Test)** – проверяет, отличается ли медиана выборки от заданного значения.
- **Критерий Колмогорова-Смирнова (Kolmogorov-Smirnov Test)** – проверяет соответствие выборки заданному распределению.

Пример t-теста в R:

```
sample_data <- c(5.2, 4.8, 5.1, 4.9, 5.3)
t.test(sample_data, mu = 5) # Проверка, отличается ли среднее от 5
```

2.2 Тесты для двух выборок

Используются для проверки различий между двумя независимыми или зависимыми выборками.

- **t-тест для независимых выборок (Independent Samples t-test)** – сравнивает средние значения двух групп.
- **U-критерий Манна-Уитни (Mann-Whitney U Test)** – непараметрический тест, сравнивающий различия между двумя независимыми группами.

- **Парный t-тест (Paired t-test)** – сравнивает значения двух зависимых выборок (до/после эксперимента).
- **Критерий Вилкоксона (Wilcoxon Signed-Rank Test)** – непараметрический аналог парного t-теста.

Пример t-теста для двух независимых выборок в R:

```
group1 <- c(5.1, 4.9, 5.0, 5.2, 5.3)
group2 <- c(4.8, 4.7, 4.9, 5.0, 4.8)
t.test(group1, group2) # Проверка различий между группами
```

Пример U-критерия Манна-Уитни:

```
wilcox.test(group1, group2)
```

Статистические тесты позволяют проверять гипотезы о данных и выявлять значимые различия между группами.

Практическая часть

1. Выполнить тест на нормальность.
2. Провести t-test и U-критерий Манна-Уитни.

Контрольные вопросы

1. Что такое статистическая гипотеза?
 2. Какие критерии используются для проверки нормальности?
 3. Чем отличается параметрический тест от непараметрического?
 4. Как интерпретировать p-value?
 5. В каких случаях применяется t-test?
-

Лабораторная работа 7. Анализ одномерных данных

Теоретическая часть

1. Методы анализа одномерных данных

Анализ одномерных данных (univariate analysis) фокусируется на изучении одной переменной и её распределения, характеристик и закономерностей. Он включает в себя расчет описательных статистик, визуализацию данных и применение статистических тестов.

1.1 Описательная статистика

- Среднее (mean) – центральное значение набора данных.
- Медиана (median) – значение, которое делит данные пополам.
- Мода (mode) – наиболее часто встречающееся значение.
- Дисперсия (variance) и стандартное отклонение (standard deviation) – степень разброса данных.
- Квартили и межквартильный размах (IQR) – для оценки распределения и выбросов.

Пример вычисления в R:

```
summary(data$column)
mean(data$column)
var(data$column)
sd(data$column)
quantile(data$column, probs = c(0.25, 0.5, 0.75))
```

1.2 Графические методы анализа

- Гистограмма (histogram) – отображает частоту значений переменной.
- Boxplot (коробчатая диаграмма) – помогает выявить выбросы и асимметрию распределения.
- Q-Q plot – проверяет соответствие распределения нормальному.

Пример построения графиков в R:

```
hist(data$column, main="Гистограмма", xlab="Значения", col="blue")
boxplot(data$column, main="Boxplot")
qqnorm(data$column)
qqline(data$column, col="red")
```

1.3 Проверка гипотез о распределении

- Тест Шапиро-Уилка (Shapiro-Wilk test) – проверяет нормальность распределения.
- Критерий Колмогорова-Смирнова (Kolmogorov-Smirnov test) – тестирует соответствие данных заданному распределению.

Пример в R:

```
shapiro.test(data$column)
ks.test(data$column, "pnorm")
```

Анализ одномерных данных позволяет глубже понять структуру данных и определить их ключевые характеристики перед проведением более сложных исследований.

2. Корреляция и регрессия

Корреляция и регрессия – это методы статистического анализа, которые используются для оценки взаимосвязи между переменными.

Корреляционный анализ показывает степень связи между переменными, а регрессионный анализ позволяет предсказать значение одной переменной на основе другой.

2.1 Корреляция

Корреляция измеряет силу и направление связи между двумя переменными. В R используется несколько коэффициентов корреляции:

- Корреляция Пирсона (Pearson) – оценивает линейную зависимость между переменными.
- Корреляция Спирмена (Spearman) – оценивает монотонную зависимость, не требующую нормального распределения данных.
- Корреляция Кендалла (Kendall) – подходит для ранговых данных.

Пример вычисления коэффициента корреляции в R:

```
correlation_pearson <- cor(data$x, data$y, method="pearson")
correlation_spearman <- cor(data$x, data$y, method="spearman")
correlation_kendall <- cor(data$x, data$y, method="kendall")
```

2.2 Регрессионный анализ

Регрессионный анализ используется для прогнозирования значений одной переменной на основе другой. Основные виды регрессии:

- Линейная регрессия – модель вида $y = a + bx$, где a – свободный член, b – коэффициент наклона.

- Множественная регрессия – учитывает несколько предикторов.
- Логистическая регрессия – применяется для предсказания категориальных переменных.

Пример построения линейной регрессии в R:

```
model <- lm(y ~ x, data=data)
summary(model)
```

2.3 Интерпретация регрессионной модели

- Коэффициент детерминации (R^2) – показывает, какая доля вариации зависимой переменной объясняется предиктором.
- p-value – указывает на значимость предиктора в модели.
- Остатки модели – разница между предсказанными и фактическими значениями, анализируются для проверки качества модели.

Графическое представление линейной регрессии в R:

```
plot(data$x, data$y)
abline(model, col="red")
```

Использование корреляции и регрессии помогает выявить зависимости в данных и строить прогнозные модели.

Практическая часть

1. Рассчитать корреляцию между переменными.
2. Построить регрессионную модель.

Контрольные вопросы

1. В чем разница между корреляцией и регрессией?
 2. Какие типы корреляции существуют?
 3. Как интерпретировать коэффициент детерминации?
 4. Какие функции используются для линейной регрессии?
 5. Как проверить значимость регрессионной модели?
-

Лабораторная работа 8. Анализ временных рядов

Теоретическая часть

1. Основы анализа временных рядов

Анализ временных рядов (Time Series Analysis) – это метод изучения данных, собранных через регулярные промежутки времени. Он применяется в экономике, финансах, метеорологии и других сферах для выявления закономерностей и прогнозирования будущих значений.

1.1 Компоненты временного ряда

- Тренд (Trend) – долгосрочное изменение уровня ряда (рост или спад).
- Сезонность (Seasonality) – периодические колебания, повторяющиеся через фиксированные промежутки времени.
- Случайные колебания (Random Noise) – нерегулярные изменения, вызванные случайными факторами.

1.2 Визуализация временных рядов

- Использование функции `plot()` для отображения временного ряда:

```
plot(ts_data, main="График временного ряда", ylab="Значение", xlab="Время",  
col="blue", type="l")
```

- Добавление тренда с помощью `abline()`:

```
abline(reg=lm(ts_data ~ time(ts_data)), col="red")
```

1.3 Декомпозиция временного ряда

Декомпозиция позволяет выделить основные компоненты временного ряда – тренд, сезонность и случайные колебания:

```
ts_decomp <- decompose(ts_data)  
plot(ts_decomp)
```

1.4 Проверка стационарности

Для применения многих методов анализа временных рядов требуется стационарность данных (отсутствие тренда и сезонности). Проверяется тестом Дики-Фуллера:

```
library(tseries)  
adf.test(ts_data)
```

Если ряд нестационарен, можно применить дифференцирование:

```
ts_diff <- diff(ts_data)
plot(ts_diff, main="Дифференцированный ряд")
```

2. Методы сглаживания

Методы сглаживания временных рядов используются для уменьшения шумов и выявления основной тенденции данных. Они помогают сгладить резкие колебания и облегчить прогнозирование.

2.1 Скользящее среднее (Moving Average, MA)

Скользящее среднее – один из наиболее простых методов сглаживания. Оно вычисляется как среднее значение фиксированного количества предыдущих наблюдений.

```
library(zoo)
ts_ma <- rollmean(ts_data, k = 5, fill = NA)
plot(ts_data, type = "l", col = "blue", main = "Скользящее среднее")
lines(ts_ma, col = "red")
```

Где k – размер окна сглаживания.

2.2 Экспоненциальное сглаживание (Exponentially Weighted Moving Average, EWMA)

Экспоненциальное сглаживание придает больше веса недавним наблюдениям, снижая влияние старых значений.

```
library(TTR)
ts_ewma <- EMA(ts_data, n = 5)
plot(ts_data, type = "l", col = "blue", main = "Экспоненциальное сглаживание")
lines(ts_ewma, col = "red")
```

Здесь n – параметр, определяющий скорость затухания.

2.3 Локально-взвешенное сглаживание (LOESS)

LOESS (Locally Estimated Scatterplot Smoothing) выполняет локальную аппроксимацию данных и хорошо подходит для нелинейных временных рядов.

```
loess_fit <- loess(ts_data ~ time(ts_data), span = 0.2)
ts_loess <- predict(loess_fit)
plot(ts_data, type = "l", col = "blue", main = "LOESS сглаживание")
lines(ts_loess, col = "red")
```

span определяет ширину окна сглаживания.

Методы сглаживания помогают устранить шумы в данных и подготовить временной ряд к прогнозированию.

Практическая часть

1. Построить временной ряд.
2. Применить методы сглаживания (МА, EWMA).

Контрольные вопросы

1. Как строить временной ряд в R?
 2. Какие функции используются для анализа тренда?
 3. Как сглаживать данные?
 4. Какие методы прогнозирования существуют?
 5. Как оценить сезонность данных?
-

Лабораторная работа 9. Визуализация данных и построение графиков

Теоретическая часть

1. Основные графические функции R

Графическое представление данных в R играет важную роль в анализе и интерпретации информации. В R встроены базовые функции визуализации, а также поддерживаются мощные библиотеки, такие как ggplot2 и lattice.

1.1 Базовые функции визуализации

R предоставляет несколько встроенных функций для создания графиков:

- `plot(x, y)`: построение точечной или линейной диаграммы.
- `hist(x)`: построение гистограммы.
- `boxplot(x)`: построение коробчатой диаграммы (boxplot).
- `barplot(x)`: построение столбчатой диаграммы.

Пример построения простого графика:

```
x <- 1:10
y <- x^2
plot(x, y, type="b", col="blue", main="График y = x^2", xlab="X", ylab="Y")
```

1.2 Использование ggplot2 для сложной визуализации

Библиотека ggplot2 предоставляет мощные средства для построения сложных визуализаций на основе концепции Grammar of Graphics.

Пример построения точечной диаграммы с ggplot2:

```
library(ggplot2)
data <- data.frame(x = 1:10, y = (1:10)^2)
ggplot(data, aes(x, y)) + geom_point(color = "red") +
  ggtitle("Точечная диаграмма")
```

1.3 Дополнительные графики

- **Гистограмма:**

```
hist(mtcars$mpg, col="lightblue", main="Гистограмма расхода топлива")
```

- **Boxplot для выявления выбросов:**

```
boxplot(mtcars$mpg, main="Boxplot расхода топлива", col="orange")
```

- **Столбчатая диаграмма:**

```
counts <- table(mtcars$cyl)
barplot(counts, main="Распределение цилиндров", col=c("red", "blue", "green"))
```

Использование этих инструментов позволяет эффективно визуализировать данные и анализировать их распределение.

2. Настройка и оформление графиков

Оформление графиков играет важную роль в представлении данных. В R можно настраивать цвета, подписи осей, легенды и другие параметры для улучшения читаемости графиков.

2.1 Настройка заголовков и подписей осей

- `main` – заголовок графика.
- `xlab` и `ylab` – подписи осей X и Y.

```
plot(x, y, main="График зависимости", xlab="Ось X", ylab="Ось Y")
```

2.2 Настройка цвета и типа линии

- `col` – цвет графика.
- `lty` – стиль линии (сплошная, пунктир и т. д.).

```
plot(x, y, type="l", col="blue", lty=2)
```

2.3 Добавление легенды

- Используется функция `legend()`.

```
legend("topright", legend=c("Линия 1", "Линия 2"), col=c("red", "blue"), lty=1:2)
```

2.4 Настройка маркеров и подписей точек

- `pch` – стиль маркеров.
- `text()` – добавление подписей к точкам.

```
plot(x, y, pch=19, col="red")
text(x, y, labels=c("A", "B", "C"), pos=3)
```

2.5 Сохранение графиков

Графики можно сохранять в файлы различных форматов.

- Сохранение в PNG:

```
png("graph.png", width=800, height=600)
plot(x, y)
dev.off()
```

- Сохранение в PDF:

```
pdf("graph.pdf")
plot(x, y)
dev.off()
```

Настройка графиков позволяет сделать визуализацию более информативной и удобной для анализа.

Практическая часть

1. Построить линейный график.
2. Настроить оформление диаграммы.

Контрольные вопросы

1. Какие пакеты используются для визуализации в R?
2. Как строить линейные и точечные диаграммы?
3. Какие параметры влияют на оформление графиков?
4. Как добавить легенду и подписи?
5. Как сохранять графики в файлы?

Эти лабораторные работы помогут студентам освоить основные методы анализа данных в R и научиться работать с реальными наборами данных.

Учебно-методическое обеспечение дисциплины

Перечень основной литературы:

1. Лонг Дж. Д. Р. Книга рецептов: проверенные рецепты для статистики, анализа и визуализации данных / Дж. Д. Лонг, П. Титор; перевод Д. А. Беликов. - Москва: ДМК Пресс, 2020. - 510 с. - ISBN 978-5-97060-835-7. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/124707.html>. — Режим доступа: для авторизир. пользователей.

Перечень дополнительной литературы:

1. Бюиссон Ф. Анализ поведенческих данных на R и Python / Ф. Бюиссон; перевод А. В. Логунов. - Москва: ДМК Пресс, 2022. - 368 с. - ISBN 978-5-97060-992-7. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/125286.html>. - Режим доступа: для авторизир. пользователей.

2. Шнарева Г. В. Анализ данных: учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. - Анализ данных. - Электрон. дан. (1 файл). - Симферополь: Университет экономики и управления, 2019. - 129 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 2227-8397, экземпляров неограниченно.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. <https://stepik.org/course/497/promo> (онлайн курс «Основы программирования на R») от Института биоинформатики)
2. <https://stepik.org/course/129/promo#toc> (онлайн курс «Анализ данных в R») от Института биоинформатики)
3. Основы программирования на R:
<https://vk.com/away.php?to=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3D94wqu3zv5o0%26list%3DPLu5flfwrnSD7wxKXFgsiuxrMKLfFHm6CD&utf=1>
4. Руководство по изучению языка R и его использование в Data Science
<https://vk.com/away.php?to=https%3A%2F%2Fproglab.io%2Fp%2Fdata-science-with-r&utf=1>
5. <http://www.intuit.ru> Национальный открытый университет ИНТУИТ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания
по организации и проведению самостоятельной работы
по дисциплине
«Программирование на R»
для студентов направления подготовки
02.03.01 Математика и компьютерные науки
направленность (профиль)
«Анализ данных и искусственный интеллект»

Общие положения

Самостоятельная работа - планируемая учебная, учебно-исследовательская, научно-исследовательская работа студентов, выполняемая во внеаудиторное (аудиторное) время по заданию и при методическом руководстве преподавателя, но без его непосредственного участия (при частичном непосредственном участии преподавателя, оставляющем ведущую роль за работой студентов).

Самостоятельная работа студентов (СРС) в ВУЗе является важным видом учебной и научной деятельности студента. Самостоятельная работа студентов играет значительную роль в рейтинговой технологии обучения.

К основным видам самостоятельной работы студентов относятся:

- формирование и усвоение содержания конспекта лекций на базе рекомендованной лектором учебной литературы, включая информационные образовательные ресурсы (электронные учебники, электронные библиотеки и др.);
- написание докладов;
- подготовка к семинарам, практическим и лабораторным работам, их оформление;
- составление аннотированного списка статей из соответствующих журналов по отраслям знаний (педагогических, психологических, методических и др.);
- выполнение учебно-исследовательских работ, проектная деятельность;
- подготовка практических разработок и рекомендаций по решению проблемной ситуации;
- выполнение домашних заданий в виде решения отдельных задач, проведения типовых расчетов, расчетно-компьютерных и индивидуальных работ по отдельным разделам содержания дисциплин и т.д.;
- компьютерный текущий самоконтроль и контроль успеваемости на базе электронных обучающих и аттестующих тестов;
- выполнение курсовых работ (проектов) в рамках дисциплин;
- выполнение выпускной квалификационной работы и др.

Методика организации самостоятельной работы студентов зависит от структуры, характера и особенностей изучаемой дисциплины, объема часов на ее изучение, вида заданий для самостоятельной работы студентов, индивидуальных качеств студентов и условий учебной деятельности.

Процесс организации самостоятельной работы студентов включает в себя следующие этапы:

- подготовительный (определение целей, составление программы, подготовка методического обеспечения, подготовка оборудования);
- основной (реализация программы, использование приемов поиска информации, усвоения, переработки, применения, передачи знаний, фиксирование результатов, самоорганизация процесса работы);
- заключительный (оценка значимости и анализ результатов, их систематизация, оценка эффективности программы и приемов работы, выводы о направлениях оптимизации труда).

Самостоятельная работа по дисциплине **«Программирование на R»** направлена на формирование следующих **компетенций**:

ПК-5 Способен принимать участие в управлении проектами создания информационных систем и программных комплексов на стадиях их жизненного цикла

ПК-7 Способен подготавливать данные и проводить аналитические работы по исследованию больших данных

1. Цель и задачи самостоятельной работы

Ведущая цель организации и осуществления СРС совпадает с целью обучения студента – формирование соответствующего набора компетенций будущего бакалавра по направлению подготовки «Математика и компьютерные науки».

При организации СРС важным и необходимым условием становятся формирование умения самостоятельной работы для приобретения знаний, навыков и возможности организации учебной и научной деятельности. Целью самостоятельной работы студентов является овладение фундаментальными знаниями, профессиональными умениями и навыками деятельности по профилю, опытом творческой, исследовательской деятельности. Самостоятельная работа студентов способствует развитию самостоятельности, ответственности и организованности, творческого подхода к решению проблем учебного и профессионального уровня.

Задачами СРС являются:

- систематизация и закрепление полученных теоретических знаний и практических умений студентов;
- углубление и расширение теоретических знаний;
- формирование умений использовать нормативную, правовую, справочную документацию и специальную литературу;
- развитие познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирование самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- развитие исследовательских умений;
- использование материала, собранного и полученного в ходе самостоятельных занятий на семинарах, на практических и лабораторных занятиях, при написании курсовых и выпускной квалификационной работ, для эффективной подготовки к итоговым зачетам и экзаменам.

2. Технологическая карта самостоятельной работы студента

Коды реализуемых компетенций	Вид деятельности студентов	Средства и технологии оценки	Объем часов, в том числе (акад.)		
			СРС	Контактная работа с преподавателем	Всего
7 семестр					
ПК-5, ПК-7	Подготовка курсовой работы	Курсовая работа	28	2	30
ПК-5, ПК-7	Решение практических задач	Конспект, отчет	22	2	24
Итого за 7 семестр			50	4	54
Итого			50	4	54

3. Порядок выполнения самостоятельной работы студентом

3.1. Методические рекомендации по работе с учебной литературой

При работе с книгой необходимо подобрать литературу, научиться правильно ее читать, вести записи. Для подбора литературы в библиотеке используются алфавитный и систематический каталоги.

Важно помнить, что рациональные навыки работы с книгой - это всегда большая экономия времени и сил.

Правильный подбор учебников рекомендуется преподавателем, читающим лекционный курс. Необходимая литература может быть также указана в методических разработках по данному курсу.

Изучая материал по учебнику, следует переходить к следующему вопросу только после правильного уяснения предыдущего, описывая на бумаге все выкладки и вычисления (в том числе те, которые в учебнике опущены или на лекции даны для самостоятельного вывода).

При изучении любой дисциплины большую и важную роль играет самостоятельная индивидуальная работа.

Особое внимание следует обратить на определение основных понятий курса. Студент должен подробно разбирать примеры, которые поясняют такие определения, и уметь строить аналогичные примеры самостоятельно. Нужно добиваться точного представления о том, что изучаешь. Полезно составлять опорные конспекты. При изучении материала по учебнику полезно в тетради (на специально отведенных полях) дополнять конспект лекций. Там же следует отмечать вопросы, выделенные студентом для консультации с преподавателем.

Выводы, полученные в результате изучения, рекомендуется в конспекте выделять, чтобы они при перечитывании записей лучше запоминались.

Опыт показывает, что многим студентам помогает составление листа опорных сигналов, содержащего важнейшие и наиболее часто употребляемые формулы и понятия. Такой лист помогает запомнить формулы, основные положения лекции, а также может служить постоянным справочником для студента.

Чтение научного текста является частью познавательной деятельности. Ее цель – извлечение из текста необходимой информации. От того на сколько осознанна читающим собственная внутренняя установка при обращении к печатному слову (найти нужные сведения, усвоить информацию полностью или частично, критически проанализировать материал и т.п.) во многом зависит эффективность осуществляемого действия.

Выделяют **четыре основные установки в чтении научного текста:**

информационно-поисковый (задача – найти, выделить искомую информацию)

усваивающая (усилия читателя направлены на то, чтобы как можно полнее осознать и запомнить как сами сведения излагаемые автором, так и всю логику его рассуждений)

аналитико-критическая (читатель стремится критически осмыслить материал, проанализировав его, определив свое отношение к нему)

творческая (создает у читателя готовность в том или ином виде – как отправной пункт для своих рассуждений, как образ для действия по аналогии и т.п. – использовать суждения автора, ход его мыслей, результат наблюдения, разработанную методику, дополнить их, подвергнуть новой проверке).

Основные виды систематизированной записи прочитанного:

Аннотирование – предельно краткое связное описание просмотренной или прочитанной книги (статьи), ее содержания, источников, характера и назначения;

Планирование – краткая логическая организация текста, раскрывающая содержание и структуру изучаемого материала;

Тезирование – лаконичное воспроизведение основных утверждений автора без привлечения фактического материала;

Цитирование – дословное выписывание из текста выдержек, извлечений, наиболее существенно отражающих ту или иную мысль автора;

Конспектирование – краткое и последовательное изложение содержания прочитанного.

Конспект – сложный способ изложения содержания книги или статьи в логической последовательности. Конспект аккумулирует в себе предыдущие виды записи, позволяет всесторонне охватить содержание книги, статьи. Поэтому умение составлять план, тезисы, делать выписки и другие записи определяет и технологию составления конспекта.

Методические рекомендации по составлению конспекта:

1. Внимательно прочитайте текст. Уточните в справочной литературе непонятные слова. При записи не забудьте вынести справочные данные на поля конспекта;
2. Выделите главное, составьте план;
3. Кратко сформулируйте основные положения текста, отметьте аргументацию автора;
4. Законспектируйте материал, четко следуя пунктам плана. При конспектировании старайтесь выразить мысль своими словами. Записи следует вести четко, ясно.
5. Грамотно записывайте цитаты. Цитируя, учитывайте лаконичность, значимость мысли.

В тексте конспекта желательно приводить не только тезисные положения, но и их доказательства. При оформлении конспекта необходимо стремиться к емкости каждого предложения. Мысли автора книги следует излагать кратко, заботясь о стиле и выразительности написанного. Число дополнительных элементов конспекта должно быть логически обоснованным, записи должны распределяться в определенной последовательности, отвечающей логической структуре произведения. Для уточнения и дополнения необходимо оставлять поля.

Овладение навыками конспектирования требует от студента целеустремленности, повседневной самостоятельной работы.

3.2. Методические рекомендации по подготовке к практическим и лабораторным занятиям

Для того чтобы практические и лабораторные занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по вычитанному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться на практических занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

3.3. Методические рекомендации по самопроверке знаний

После изучения определенной темы по записям в конспекте и учебнику, а также решения достаточного количества соответствующих задач на практических занятиях и самостоятельно студенту рекомендуется, провести самопроверку усвоенных знаний, ответив на контрольные вопросы по изученной теме.

В случае необходимости нужно еще раз внимательно разобраться в материале.

Иногда недостаточность усвоения того или иного вопроса выясняется только при изучении дальнейшего материала. В этом случае надо вернуться назад и повторить плохо усвоенный материал. Важный критерий усвоения теоретического материала - умение решать задачи или пройти тестирование по пройденному материалу. Однако следует помнить, что правильное решение задачи может получиться в результате применения механически заученных формул без понимания сущности теоретических положений.

3.4. Методические рекомендации по написанию научных текстов (докладов, рефератов, эссе, научных статей и т.д.)

Перед тем, как приступить к написанию научного текста, важно разобраться, какова истинная цель вашего научного текста - это поможет вам разумно распределить свои силы и время.

Во-первых, сначала нужно определиться с идеей научного текста, а для этого необходимо научиться либо относиться к разным явлениям и фактам несколько критически (своя идея – как иная точка зрения), либо научиться увлекаться какими-то известными идеями, которые нуждаются в доработке (идея – как оптимистическая позиция и направленность на дальнейшее совершенствование уже известного). Во-вторых, научиться организовывать свое время, ведь, как известно, свободное (от всяких глупостей) время – важнейшее условие настоящего творчества, для него наконец-то появляется время. Иногда именно на организацию такого времени уходит немалая часть сил и талантов.

Писать следует ясно и понятно, стараясь основные положения формулировать четко и недвусмысленно (чтобы и самому понятно было), а также стремясь структурировать свой текст. Каждый раз надо представлять, что ваш текст будет кто-то читать и ему захочется сориентироваться в нем, быстро находить ответы на интересующие вопросы (заодно представьте себя на месте такого человека). Понятно, что работа, написанная «сплошным текстом» (без заголовков, без выделения крупным шрифтом наиболее важным мест и т. п.), у культурного читателя должна вызывать брезгливость и даже жалость к автору (исключения составляют некоторые древние тексты, когда и жанр был иной и к текстам относились иначе, да и самих текстов было гораздо меньше – не то, что в эпоху «информационного взрыва» и соответствующего «информационного мусора»).

Объем текста и различные оформительские требования во многом зависят от принятых в конкретном учебном заведении порядков.

Реферат (доклад) - это самостоятельное исследование студентом определенной проблемы, комплекса взаимосвязанных вопросов.

Реферат не должна составляться из фрагментов статей, монографий, пособий. Кроме простого изложения фактов и цитат, в реферате должно проявляться авторское видение проблемы и ее решения.

Рассмотрим основные этапы подготовки реферата студентом.

Выполнение реферата начинается с выбора темы.

Затем студент приходит на первую консультацию к руководителю, которая предусматривает:

- обсуждение цели и задач работы, основных моментов избранной темы;
- консультирование по вопросам подбора литературы;
- составление предварительного плана.

Следующим этапом является работа с литературой. Необходимая литература подбирается студентом самостоятельно.

После подбора литературы целесообразно сделать рабочий вариант плана работы. В нем нужно выделить основные вопросы темы и параграфы, раскрывающие их содержание.

Составленный список литературы и предварительный вариант плана уточняются, согласуются на очередной консультации с руководителем.

Затем начинается следующий этап работы - изучение литературы. Только внимательно читая и конспектируя литературу, можно разобраться в основных вопросах темы и подготовиться к самостоятельному (авторскому) изложению содержания реферата. Конспектируя первоисточники, необходимо отразить основную идею автора и его позицию по исследуемому вопросу, выявить проблемы и наметить задачи для дальнейшего изучения данных проблем.

Систематизация и анализ изученной литературы по проблеме исследования позволяют студенту написать работу.

Рабочий вариант текста реферата предоставляется руководителю на проверку. На основе рабочего варианта текста руководитель вместе со студентом обсуждает возможности доработки текста, его оформление. После доработки реферат сдается на кафедру для его оценивания руководителем.

Требования к написанию реферата

Написание 1 реферата является обязательным условием выполнения плана СРС по любой дисциплине профессионального цикла.

Тема реферата может быть выбрана студентом из предложенных в рабочей программе или фонде оценочных средств дисциплины, либо определена самостоятельно, исходя из интересов студента (в рамках изучаемой дисциплины). Выбранную тему необходимо согласовать с преподавателем.

Реферат должен быть написан научным языком.

Объем реферата должен составлять 20-25 стр.

Структура реферата:

- Введение (не более 3-4 страниц). Во введении необходимо обосновать выбор темы, ее актуальность, очертить область исследования, объект исследования, основные цели и задачи исследования.

- Основная часть состоит из 2-3 разделов. В них раскрывается суть исследуемой проблемы, проводится обзор мировой литературы и источников Интернет по предмету исследования, в котором дается характеристика степени разработанности проблемы и авторская аналитическая оценка основных теоретических подходов к ее решению. Изложение материала не должно ограничиваться лишь описательным подходом к раскрытию выбранной темы. Оно также должно содержать собственное видение

рассматриваемой проблемы и изложение собственной точки зрения на возможные пути ее решения.

- Заключение (1-2 страницы). В заключении кратко излагаются достигнутые при изучении проблемы цели, перспективы развития исследуемого вопроса

- Список использованной литературы (не меньше 10 источников), в алфавитном порядке, оформленный в соответствии с принятыми правилами. В список использованной литературы рекомендуется включать работы отечественных и зарубежных авторов, в том числе статьи, опубликованные в научных журналах в течение последних 3-х лет и ссылки на ресурсы сети Интернет.

- Приложение (при необходимости).

Требования к оформлению:

- текст с одной стороны листа;
- шрифт Times New Roman;
- кегль шрифта 14;
- межстрочное расстояние 1,5;
- поля: сверху 2,5 см, снизу – 2,5 см, слева - 3 см, справа 1,5 см;
- реферат должен быть представлен в сброшюрованном виде.

Порядок защиты реферата:

Защита реферата проводится на практических занятиях, после окончания работы студента над ним и исправления всех недочетов, выявленных преподавателем в ходе консультаций. На защиту реферата отводится 5-7 минут времени, в ходе которого студент должен показать свободное владение материалом по заявленной теме. При защите реферата приветствуется использование мультимедиа-презентации.

Оценка реферата

Реферат оценивается по следующим критериям:

- соблюдение требований к его оформлению;
- необходимость и достаточность для раскрытия темы приведенной в тексте реферата информации;
- умение студента свободно излагать основные идеи, отраженные в реферате;
- способность студента понять суть задаваемых преподавателем и сокурсниками вопросов и сформулировать точные ответы на них.

Критерии оценки:

Оценка «отлично» выставляется студенту, если в докладе студент исчерпывающе, последовательно, четко и логически стройно излагает материал; свободно справляется с задачами, вопросами и другими видами применения знаний; использует для написания доклада современные научные материалы; анализирует полученную информацию; проявляет самостоятельность при написании доклада.

Оценка «хорошо» выставляется студенту, если качество выполнения доклада достаточно высокое. Студент твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопросы по теме доклада.

Оценка «удовлетворительно» выставляется студенту, если материал доклада излагается частично, но пробелы не носят существенного характера, студент допускает неточности и ошибки при защите доклада, дает недостаточно правильные формулировки, наблюдаются нарушения логической последовательности в изложении материала.

Оценка «неудовлетворительно» выставляется студенту, если он не подготовил доклад или допустил существенные ошибки. Студент неуверенно излагает материал доклада, не отвечает на вопросы преподавателя.

Описание шкалы оценивания

Максимально возможный балл за весь текущий контроль устанавливается равным 55. Текущее контрольное мероприятие считается сданным, если студент получил за него не менее 60% от установленного для этого контроля максимального балла. Рейтинговый балл, выставляемый студенту за текущее контрольное мероприятие, сданное студентом в установленные графиком контрольных мероприятий сроки, определяется следующим образом:

Уровень выполнения контрольного задания	Рейтинговый балл (в % от максимального балла за контрольное задание)
Отличный	100
Хороший	80
Удовлетворительный	60
Неудовлетворительный	0

Темы рефератов по дисциплине «Программирование на R»

1. Функции, связанные с распределениями вероятностей.
2. Возможности графики. Использование аргументов графических команд.
3. Возможности графики. Использование команды *par()*.
4. Варианты команды *apply()*: *sapply()*, *tapply()*. А также *mapply()* (сложное).
5. Команда *aggregate()*.
6. Функции, работающие с символьными данными.
7. Упорядоченный фактор: создание, использование.
8. Команда *outer()*. Трехмерные графики функций.
9. Представление даты и времени. Функции работы с датами.
10. Временные ряды. Визуализация, первичная обработка.
11. Функция *summary()*, варианты использования (в том числе для линейных моделей).
12. Формулы моделей. Использование в проверках гипотез и для линейных моделей.
13. Обобщенная линейная модель. Функция *glm()*.
14. Оценка мощности критерия. Команды подбора параметров по мощности.
15. Сравнение нескольких выборок. Тесты типа *pairwise.**
16. Иерархическая кластеризация: варианты.
17. Обработка пропущенных значений.
18. Графический пакет *lattice*.
19. Графический пакет *ggplot2*.
20. Сравните с помощью расчетов и моделирования две статистики для гипотезы о корреляции (преобразование Фишера и статистику

Учебно-методическое обеспечение СРС

Перечень основной литературы:

1. Лонг Дж. Д. Р. Книга рецептов: проверенные рецепты для статистики, анализа и визуализации данных / Дж. Д. Лонг, П. Титор; перевод Д. А. Беликов. - Москва: ДМК Пресс, 2020. - 510 с. - ISBN 978-5-97060-835-7. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/124707.html>. — Режим доступа: для авторизир. пользователей.

Перечень дополнительной литературы:

1. Бюиссон Ф. Анализ поведенческих данных на R и Python / Ф. Бюиссон; перевод А. В. Логунов. - Москва: ДМК Пресс, 2022. - 368 с. - ISBN 978-5-97060-992-7. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/125286.html>. - Режим доступа: для авторизир. пользователей.
2. Шнарева Г. В. Анализ данных: учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. - Анализ данных. - Электрон. дан. (1 файл). - Симферополь: Университет экономики и управления, 2019. - 129 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 2227-8397, экземпляров неограниченно.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. <https://stepik.org/course/497/promo> (онлайн курс «Основы программирования на R») от Института биоинформатики)
2. <https://stepik.org/course/129/promo#toc> (онлайн курс «Анализ данных в R») от Института биоинформатики)
3. Основы программирования на R:
<https://vk.com/away.php?to=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3D94wqu3zv5o0%26list%3DPLu5flfwrnSD7wxKXFgsiuxrMKLfFHm6CD&utf=1>
4. Руководство по изучению языка R и его использование в Data Science
<https://vk.com/away.php?to=https%3A%2F%2Fproglab.io%2Fp%2Fdata-science-with-r&utf=1>
5. <http://www.intuit.ru> Национальный открытый университет ИНТУИТ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания
по выполнению курсовой работы по дисциплине
«Программирование на R»
для студентов направления подготовки
02.03.01 Математика и компьютерные науки

Введение

Настоящие рекомендации предназначены для студентов всех форм обучения при выполнении курсовых работ, для руководителей курсовых работ с целью формирования единых требований при разработке и защите работ. Рекомендации содержат предложения по структуре курсовой работы (далее - работа), объему, содержанию, оформлению, процедуре допуска к защите и порядку защиты работ.

Целями выполнения курсовой работы являются:

- обогащение знаний студентов,
- обучение методам теоретического анализа явлений и закономерностей науки,
- выработка навыков применения теоретических знаний к комплексному решению профессиональных задач, использования справочной литературы, методов математической обработки экспериментальных данных, компьютерных технологий.

Системой курсовых работ (проектов) студент подготавливается к выполнению выпускной квалификационной работы.

Перед прочтением Методических указаний ознакомьтесь с **основными ошибками**, которые допускают студенты, приступая к написанию курсовой работы. Возможно, это мотивирует на более внимательное отношение к требованиям по написанию курсовой работы.

1. Выбрав тему, студент выходит в сеть Интернет и через поисковую систему (Yandex, Google и т.д. – зависит от предпочтений) находит курсовую работу по данной теме, меняет реквизиты, Ф.И.О. автора и сдает преподавателю. В результате получает рецензию: «Плагат. На переработку».
2. Аналогичные действия п.1, но студент подходит к решению проблемы творчески и находит несколько курсовых работ по данной теме. Компонует материал по своему усмотрению и сдает преподавателю под своим авторством. В результате получает рецензию: «Плагат. На переработку». Почему дана такая рецензия? По трем основным причинам: во-первых, написание курсовых работ – распространённый бизнес в Интернете и качественный материал бесплатно не выкладывают в сети, таким образом, компоновать приходится из «третьесортного сырья»; во-вторых, курсовая уже написана и защищена ранее, то есть она уже устарела и не отражает современных реалий, в-третьих, такие работы редко соответствуют требованиям, предъявляемым кафедрой к курсовым работам.
3. Студент принимает постороннюю помощь в написании работы. Здесь есть несколько аспектов, но следует обратить внимание на два основных. Во-первых, помощь исходит, как правило, от одного физического или юридического лица для группы студентов. В результате преподаватель получает на проверку несколько работ на разные темы, написанные «одной рукой». Формальных причин для отрицательной рецензии у преподавателя нет, особенно если работа выполнена в строгом соответствии с требованиями, но, как показывает практика, защита таких работ проходит безуспешно. Во-вторых, студент, не приложивший усилий к данной работе, как правило, очень плохо в ней ориентируется, особенно, когда преподаватель уточняет аспекты рассматриваемой темы. Но так как Вы уже вышли на защиту, то получаете заслуженную оценку «неудовлетворительно».
4. Вы самостоятельно осуществляете работу, но так же самостоятельно выбираете требования к курсовой работе. Если Вам они нравятся и понятны, то исполняете, а если не нравятся, пропускаете. Это очень распространённая ошибка. **Выполнять надо все без исключения требования к курсовой работе.** Если в процессе работы возникли вопросы, следует по ним получить консультацию у научного руководителя.

Это краткий перечень основных ошибок, которые недопустимо совершать после прочтения данных Методических указаний.

1. ОРГАНИЗАЦИЯ ВЫПОЛНЕНИЯ И ЗАЩИТЫ КУРСОВОЙ РАБОТЫ

Виды работ по этапам выполнения курсовой работы:

а) Предварительный этап:

- выбор темы курсовой работы и оценка возможности раскрытия данной темы;
- подача заявления на закрепление темы;

б) Основной этап:

- сбор материала;
- оформление курсовой работы;

в) Заключительный этап:

- рецензирование работы руководителем и допуск к защите;
- защита работы.

1.1 Предварительный этап

Темы курсовых работ определяются выпускающей кафедрой. Студенту предоставлено право выбора темы курсовой работы. Задание к выбранной теме выдается руководителем работы (проекта) и регистрируется в кафедральном журнале.

1.2 Основной этап

Составление плана курсовой работы

План курсовой работы представляет собой составленный в определенном порядке перечень глав и развернутый перечень вопросов, которые должны быть освещены. Правильно составленный план работы помогает систематизировать материал, обеспечивает последовательность его изложения. План работы составляется самостоятельно, с учетом замысла и индивидуального подхода. План работы рекомендуется **согласовать с руководителем курсовой работы**. В процессе работы план работы может уточняться.

Подбор литературы

Курсовая работа выполняется на основе глубокого изучения литературных источников. Литература по теме работы может быть подобрана студентом при помощи предметных и алфавитных каталогов библиотек. Для этих целей могут быть использованы каталоги книг, указатели журнальных статей, специальные библиографические справочники, тематические сборники литературы, периодически выпускаемые отдельными издательствами, интернет-сайты официальных организаций.

Проработка источников сопровождается выписками, конспектированием. Выписки из текста делают обычно в виде цитаты. После каждой цитаты приводится ссылка на автора и источник. Поэтому при выписке цитат и конспектировании следует делать ссылки: автор, название издания, место издания, издательство, год издания, номер страницы. Конспект может содержать в себе факты, доказательства, примеры, основные положения и выводы автора, а также личное отношение студента к изученному материалу.

Рациональной и эффективной организации исследовательской деятельности способствует составление студентом собственной картотеки проработанных по теме и смежным вопросам литературных источников. Картотека не только помогает получить представление о степени изученности данной проблемы, но и позволяет работать с литературой более целеустремленно.

Оформление курсовой работы

Оформление проекта или работы осуществляется в соответствии с требованиями положения о выполнении и защите курсовых о выполнении и защите курсовых работ (проектов).

Содержание и структура курсовых работ

Курсовая работа должна содержать элементы новизны, наряду с фундаментальным аспектом должен быть проведен анализ современного состояния изучаемой проблемы, а

также отражены региональные особенности. Задание по курсовой работе необходимо индивидуализировать с учетом интересов и способностей студентов.

Курсовая работа должна состоять из введения, теоретической части, эмпирической (практической, расчетно-графической) части, заключения, списка литературы и приложения. В отдельных случаях, в соответствии с тематикой работы, эмпирическая часть может отсутствовать.

Во введении обосновывается актуальность выбранной темы исследования, задачи, цели, новизна, теоретическая и практическая значимость исследования.

Теоретическая часть должна содержать анализ состояния изучаемой проблемы на основе обзора научной, научно-информационной, справочной литературы. Представленный материал должен быть логически связан с целью исследования. В параграфах теоретической части необходимо отражать отдельные компоненты проблемы и завершать их выводами.

Эмпирическая (практическая, расчетно-графическая) часть (при наличии) включает описание системы экспериментального исследования, обоснование методов исследования, анализ результатов экспериментального исследования, схемы, графические и математические способы интерпретации полученных данных, выводы.

Заключение содержит выводы, подтверждающие или опровергающие первоначальные предположения (гипотезы), перспективы дальнейшего изучения проблемы, связь с практикой, анализ реализации целей и задач исследования.

Список литературы должен быть составлен в соответствии с требованиями к оформлению библиографий.

Приложение содержит весь фактический материал экспериментальных исследований (анкеты, опросники, схемы, чертежи, расчетные материалы, карты, рисунки, ответы респондентов и т.д.).

Содержание курсовой работы (проекта) рекомендуется представлять в объеме 20-30 страниц печатного текста. Текст работы должен быть напечатан через 1,5 интервала на одной стороне стандартного листа белой бумаги (А - 4). Текст и другие отпечатанные элементы работы должны быть черными, контуры букв и знаков – четкими, без ореола и затенения. Шрифт TimesNewRoman, кегель 14. Курсив и подчеркивание в работе не допускаются. Названия глав и параграфов выделяются полужирным шрифтом. Лист с текстом должен иметь поля: слева – 30 мм, справа – 10 мм, сверху – 20 мм, снизу – 20 мм. Нумерация страниц текста делается в правом нижнем углу листа. Проставлять номер страницы необходимо со страницы, где печатается "Введение", на которой ставится цифра "3". После этого нумеруются все страницы, включая приложения.

Между названием главы и названием параграфа этой главы ставится пробел равный двум интервалам, а название параграфа не должно отделяться от текста этого параграфа пробелом. Названия параграфов отделяются от текста предыдущего параграфа пробелом, равным двум интервалам. Каждая глава, а также введение, выводы, предложения и список использованной литературы начинаются с новой страницы. Слово "Глава" не пишется. Главы имеют порядковые номера в пределах всей работы, обозначаемые арабскими цифрами (например: 1,2,3), после которых ставится точка. Слово "параграф" или значок параграфа в названии не ставятся. Параграфы имеют порядковые номера в пределах глав, обозначаемые арабскими цифрами (например: 1.1. и 1.2.). Заголовки глав и параграфов в тексте работы должны располагаться по центру, точку в конце названия главы и параграфа не ставят. Не допускается переносить часть слова в заголовке.

Нумерация таблиц и рисунков может быть сквозной или соотноситься с номером главы и параграфа. Например, если таблица или рисунок включены в текст первого параграфа второй главы, нумерация следующая: Таблица 2.1.1., рис. 2.1.1. Последняя цифра означает порядковый номер таблицы (или рисунка) в данном параграфе. Таблица помещается в качестве следующей страницы после первого упоминания о ней в тексте.

В рамках отведенного для руководства курсовыми работами времени регулярно проводятся индивидуальные консультации. Руководитель работы осуществляет предварительный просмотр выполненных заданий. После проверки выполнения студентом одного этапа работы, руководитель работы разрешает студенту перейти к следующему.

1.3 Заключительный этап

Работа сдается руководителю. Если работа удовлетворяет требованиям, предъявляемым к курсовым работам, она допускается к защите.

Защита курсовой работы является обязательной формой проверки выполнения работы. Защита производится на заседании кафедры, научно-методического семинара кафедры, научной проблемной группы при непосредственном участии руководителя, в присутствии студентов. Результаты наиболее интересных курсовых работ могут быть доложены на научных конференциях.

Защита состоит в коротком докладе студента по выполненной работе и в ответах на вопросы присутствующих на защите.

Результаты защиты курсовой работы, согласно действующему Положению о текущем контроле и промежуточной аттестации, оцениваются дифференцированной отметкой по пятибалльной системе.

Студент, не представивший в установленный срок работу, получивший неудовлетворительную оценку на защите или не явившийся на защиту по неуважительной причине, считается имеющим академическую задолженность.

Для ликвидации академической задолженности студент обязан в установленные сроки представить курсовую работу руководителю и защитить её перед комиссией.

Допуск к защите курсовой работы

Допуск к защите осуществляет научный руководитель курсовой работы.

Работа не допускается к защите, если она не носит самостоятельного характера, списана из литературных источников или у других авторов, если основные вопросы не раскрыты, изложены схематично, фрагментарно, в тексте содержатся ошибки, научный аппарат оформлен неправильно, текст написан небрежно.

Неудовлетворительно выполненная работа подлежит переработке в соответствии с замечаниями преподавателя, содержащимися в отзыве.

При получении допуска к защите студент вправе получить дополнительную консультацию по предстоящей защите.

Курсовая работа, студенту не возвращается и хранится на кафедре в течение 2-х лет.

После получения допуска к защите, студент самостоятельно готовится к защите: составляет текст доклада, реагирует на замечания руководителя.

Подготовка к защите курсовой работы включает устранение ошибок и недостатков, изучение дополнительных источников, готовность объяснить любые приведенные в работе положения.

В ходе защиты курсовой работы задача студента - показать углубленное понимание вопросов конкретной темы, хорошее владение материалом по теме.

На защиту курсовой работы студентом предоставляются:

- курсовая работа,
- зачётная книжка.

При отсутствии одного пункта из данного перечня защита курсовой работы не может быть осуществлена.

Защита работы производится в соответствии с планом приёма защит курсовых работ. С ним можно ознакомиться на стенде кафедры или непосредственно уточнить у научного руководителя время и место защиты.

По результатам защиты курсовой работы в зачётную ведомость (для защиты курсовой работы) выставляется оценка.

Критерии оценки курсовой работы:

- **«отлично»** выставляется студенту, показавшему глубокие знания, которые применяются при самостоятельном исследовании избранной темы, умение обобщать практический материал и делать на основе анализа выводы. Курсовая работа с оценкой «отлично» может быть рекомендована на конкурс студенческих работ, использована в качестве доклада на научно-студенческой конференции.
- **«хорошо»** выставляется студенту, показавшему в работе и при её защите полное знание материала, всесторонне осветившему вопросы темы, но не полной мере проявившему самостоятельность в исследовании.
- **«удовлетворительно»** выставляется студенту, раскрывшему в работе основные вопросы избранной темы, но не проявившему самостоятельности в анализе или допустившему отдельные неточности содержания работы.
- **«неудовлетворительно»** выставляется студенту, не раскрывшему основные положения избранной темы и допустившему грубые ошибки в содержании работы.

2. СОДЕРЖАНИЕ КУРСОВОЙ РАБОТЫ

2.1 Структурные элементы курсовой работы

Структурными элементами курсовой работы являются:

- 1) титульный лист,
- 2) оглавление,
- 5) введение;
- 6) основная часть;
- 7) заключение;
- 8) список использованных источников;
- 9) приложения.

Титульный лист является первой страницей курсовой работы.

Оглавление включает введение, наименование всех глав и параграфов, заключение, список использованных источников и приложения с указанием номеров страниц, с которых начинаются эти элементы работы. Следует обратить внимание, что через оглавление прослеживается структура всей работы. Желательно основную часть работы уместить в 2 главы, каждая из которых разбивается на 2-3 параграфа. Глава не может содержать один параграф. Наличие в главе более трех параграфов говорит о поверхностном представлении данных вопросов, так как к работе предъявляются жесткие требования по объему работы.

Введение. Во введении обосновывается **актуальность темы, формулируются цель** курсовой работы **и комплекс взаимосвязанных задач**, подлежащих решению в процессе работы, а также представляется **теоретическая и методологическая база**.

Актуальность темы отражает степень важности её раскрытия на данный момент, то есть, почему читателю будет интересно, важно или просто необходимо познакомиться с Вашей работой. Это 2-3 предложения, которые анонсируют Вашу работу. Важно во введении, раскрывая актуальность темы, не уйти в раскрытие заявленной проблемы. Не следует писать фразы типа: «Мне эта тема очень интересна, поэтому она актуальна». Здесь необходимо заинтересовать читателя предметно и содержательно, а не своим примером.

Цель работы – это результат, к которому Вы желаете прийти и путь к нему. Цель работы одна. Формулируем её с использованием неопределённой формы глагола: изучить, исследовать, проанализировать, рассмотреть и т.д. Цель вашей курсовой работы скрыта в названии работы.

Задачи работы отражают шаги, делая которые, Вы достигаете цели. Вы должны поставить 6 - 7 задач. Они фактически связаны с параграфами Вашей работы. Задачи

формулируются аналогично цели с использованием неопределенной формы глагола: изучить, исследовать, проанализировать, рассмотреть, оценить, выявить и т.д.

Теоретическая и методологическая база подразумевает краткое описание использованных источников с указанием основных авторов. Это «поклон» в адрес людей, чьими трудами Вы воспользовались, а также возможность для специалиста с первого взгляда понять, о чем написана работа и какого подхода к данной проблеме придерживаетесь Вы.

Объем этой части работы 1 лист.

Так как введение – это та часть работы, которая подразумевает 100% самостоятельное изложение, важно не перейти на «ты» в общении с подразумеваемым читателем, то есть **всё изложение материала должно осуществляться в неопределенной форме глагола**. Нельзя употреблять местоимение «я».

Основная часть. Требования к содержанию основной части работы даны ранее.

Заключение. В заключении формулируются краткие выводы, полученные в ходе выполнения поставленных задач и цели. Вы кратко излагаете то, что написали в работе, но это самое главное, что вы хотите донести до читателя, самый «сок» Вашего труда. В заключение можно вынести числовые данные в целях иллюстрации вывода. Они должны отражать цель и задачи исследования, и не быть «вырванными» из работы числами.

Список использованных источников должен содержать сведения об источниках, использованных при выполнении работы.

Приложения. В приложении рекомендуется включать материалы, связанные с выполненной работой, которые по каким-либо признакам не могут быть включены в основную часть.

2.2 Содержание основной части курсовой работы

Теоретический раздел

Раздел должен иметь наименование, отражающее теоретические аспекты темы курсовой работы. В общем случае в этом разделе должны быть отражены:

1) характеристика предмета исследования в соответствии темой курсовой работы, в частности, должен присутствовать анализ понятийной базы по заявленной теме.

2) анализ состояния вопроса, являющегося темой работы, при необходимости – методы анализа.

3) основные подходы к решению проблемы, являющейся темой курсовой работы.

Теоретический раздел должен давать ответы на вопросы, что представляет собой проблема, являющаяся темой курсовой работы.

Объем раздела – 10 - 12 с. В разделе должно быть **не менее двух и не более трех подразделов**.

Аналитический раздел

Наименование данного раздела должно быть связано с анализом состояния предмета исследования, являющегося темой курсовой работы.

Исследование проблемы должно быть целевым и глубоким.

В аналитическом разделе используются различные приемы анализа, указываются источники информации, приводятся аналитические таблицы и расчеты, графические способы отражения результатов анализа, делаются выводы.

В данном разделе обязательно использование не только статистического материала, но и результатов анализа, проведенного специалистами по данному вопросу, которые изложены в источниках периодической печати.

Объем аналитического раздела курсовой работы не должен превышать 10-12 страниц.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Перечень основной литературы:

1. Лонг Дж. Д. Р. Книга рецептов: проверенные рецепты для статистики, анализа и визуализации данных / Дж. Д. Лонг, П. Титор; перевод Д. А. Беликов. - Москва: ДМК Пресс, 2020. - 510 с. - ISBN 978-5-97060-835-7. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/124707.html>. — Режим доступа: для авторизир. пользователей.

Перечень дополнительной литературы:

1. Бюиссон Ф. Анализ поведенческих данных на R и Python / Ф. Бюиссон; перевод А. В. Логунов. - Москва: ДМК Пресс, 2022. - 368 с. - ISBN 978-5-97060-992-7. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/125286.html>. - Режим доступа: для авторизир. пользователей.
2. Шнарева Г. В. Анализ данных: учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. - Анализ данных. - Электрон. дан. (1 файл). - Симферополь: Университет экономики и управления, 2019. - 129 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 2227-8397, экземпляров неограниченно.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. <https://stepik.org/course/497/promo> (онлайн курс «Основы программирования на R») от Института биоинформатики)
2. <https://stepik.org/course/129/promo#toc>(онлайн курс «Анализ данных вR») от Института биоинформатики)
3. Основы программирования на R:
<https://vk.com/away.php?to=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3D94wqu3zv5o0%26list%3DPLu5flfwrnSD7wxKXFgsiuxrMKLfFHm6CD&utf=1>
4. Руководство по изучению языка R и его использование в Data Science
<https://vk.com/away.php?to=https%3A%2F%2Fproglab.io%2Fp%2Fdata-science-with-r&utf=1>
5. <http://www.intuit.ru> Национальный открытый университет ИНТУИТ