

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Андрей Драгомирович Хлутков  
Должность: директор  
Дата подписания: 02.12.2024 23:48:09  
Уникальный программный ключ:  
880f7c07c583b07b775f6604a630281b13ca9fd2

Примерная форма  
рабочей программы дисциплины  
(с применением ЭК)

**Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И  
ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ  
ФЕДЕРАЦИИ»**

**Северо-Западный институт управления, факультет Экономики и Финансов,  
Кафедра Бизнес-информатики  
(наименование структурного подразделения (института/факультета/филиала))**

УТВЕРЖДЕНО

Директор СЗИУ РАНХиГС  
А.Д.Хлутков

**ПРОГРАММА БАКАЛАВРИАТА  
«Бизнес-аналитика»**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Б1.В.17 Язык программирования R  
(код и наименование РПД)**

**38.03.05 Бизнес-информатика  
(код, наименование направления подготовки/специальности)**

**очная  
(форма(формы) обучения)**

Год набора - 2024 г.

Санкт-Петербург, 2024 г.

**Авторы-составители РПД:**

Кандидат технических наук, доцент, доцент кафедры бизнес-информатики  
Полянская Светлана Владимировна  
Ст. преподаватель кафедры бизнес-информатики  
Веприкова Виктория Федоровна

**Заведующий кафедрой** бизнес-информатики, доктор военных наук, профессор  
Наумов Владимир Николаевич

РПД одобрена на заседании кафедры бизнес-информатики. Протокол от  
27.06.2024 г. № 10

## СОДЕРЖАНИЕ

1.	Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы .....	4
2.	Объем и место дисциплины в структуре образовательной программы..	6
3.	Содержание и структура дисциплины.....	6
4.	28. Команды кластеризации (k-means(), hclust(), cutree(...)).....	9
5.	Оценочные материалы промежуточной аттестации по дисциплине....	13
6.	Методические материалы для освоения дисциплины.....	17
7.	Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет" ,	17
8.	Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы .....	19

**1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения программы**

1.1. Дисциплина Б1.В.17 Язык программирования R обеспечивает овладение следующими компетенциями:

Таблица 1.1

Код компетенции	Наименование компетенции	Код компонента компетенции	Наименование компонента компетенции
ПКС 5	Способен решать задачи анализа больших данных с использованием существующих программных средств и технологий	ПКС 5.1	Знает модели данных, основные методы получения, хранения и обработки информации
		ПКС 5.2	Умеет анализировать, структурировать информацию и разрабатывать базы данных информационных систем

1.2. В результате освоения дисциплины у студентов должны быть сформированы:

Таблица 1.2

ОТФ/ТФ (при наличии профстандарта)/ профессиональные действия	Код компонента компетенции	Результаты обучения
Знание модели данных, основные методы получения, хранения и обработки информации	ПКС-5.1	<b>на уровне знаний:</b> - научные методы обработки и визуализации данных
		<b>на уровне умений:</b> - визуализировать, интерпретировать и давать рекомендации на основании результатов интеллектуального анализа больших данных
		<b>на уровне навыков:</b> - способы хранения, поиска, извлечения и обработки больших данных
Анализ, обоснование и выбор решения	ПКС-5.2	<b>на уровне знаний:</b> - набор программных средств для работы с данными, вычислений и графического отображения
		<b>на уровне умений:</b> - разработки новых методов интерактивного анализа данных
		<b>на уровне навыков:</b>

		- решения профессиональных задач посредством инструментов языка программирования R
--	--	------------------------------------------------------------------------------------

## 2. Объем и место дисциплины в структуре ОП ВО

### Объем дисциплины

Общая трудоемкость дисциплины составляет 2 зачетных единицы 72/54 академ. часов. Дисциплина осваивается с применением электронного (онлайн) курса.

Таблица 2.1

Вид работы	Трудоемкость (акад/астр.часы)
<b>Общая трудоемкость</b>	72/54
<b>Контактная работа с преподавателем</b>	26/19,5
Лекции	
Практические занятия	24/18
Лабораторные занятия	
<b>Практическая подготовка</b>	
<b>Самостоятельная работа</b>	46/34,5
Контроль	
Формы текущего контроля	О/КЗ
<b>Форма промежуточной аттестации</b>	Зачет с оценкой

### Место дисциплины в структуре ОП ВО

Дисциплина Б1.В.16 «Язык программирования R» входит в вариативную часть дисциплин учебного плана по направлению подготовки бакалавриата «Бизнес-информатика» 38.03.05. Преподавание дисциплины «Язык программирования R» основано на дисциплинах –Б1.О.11 «Программирование, Б1.В.15 «Введение в науку о данных. SQL и Python», Б1.В.16 «Системы обработки и анализа данных». В свою очередь она создаёт необходимые предпосылки для освоения программ таких дисциплин, как Б1.О.07.05 «Теория вероятностей и математическая статистика», Б1.В.06 «Анализ данных», Б1.В.ДВ.03.01 «Методы прогнозирования», Б1.В.ДВ.03.02 «Прогнозирование временных рядов» и ряда дисциплин по выбору студента.

Дисциплина изучается в 3-м семестре 2-го курса.

Формой промежуточной аттестации в соответствии с учебным планом является зачет.

Доступ к системе дистанционных образовательных технологий осуществляется каждым обучающимся самостоятельно. Пароль и логин к личному кабинету / профилю предоставляется студенту в деканате.

## 3. Содержание и структура дисциплины

Таблица 3.1

№ п/п	Наименование тем	Объем дисциплины, час.			Форма текущего контроля
		Всего	Контактная работа обучающихся с	СР	

			преподавателем по видам учебных занятий				успеваемос- ти**, промежуто- чной аттестации ***	
			Л	ЛР	ПЗ	КСР		
Тема 1	Общее представление о задачах и особенностях R.	10			4		6	О
Тема 2	Классы объектов, математические и статистические функции преобразования данных	12			4		8	КЗ
Тема 3	Основы управления данными	12			4		8	КЗ
Тема 4	Управляющие конструкции в R. Циклы for и while. Функции в R.	12			4		8	КЗ
Тема 5	Визуализация данных.	12			4		8	КЗ
Тема 6	Элементы корреляционного, регрессионного, кластерного анализа.	12			4		8	КЗ
Контроль								
<b>Промежуточная аттестация</b>						2		Зачет с оценкой
Всего (акад./астр. часы):		72/54			24/18		46/34,5	

Консультация к экзамену – 2 часа

КЗ – контрольное задание;

О – опрос

### Содержание дисциплины

#### *Тема 1. Общее представление о задачах и особенностях R*

Установка R и RStudio. Консоль R. Знакомство с интерфейсом RStudio. Установка загрузка библиотек в R. Знакомство с документацией R Ввод данных из разных источников. Вывод результатов расчета в файл

*Тема 2. Классы объектов, математические и статистические функции преобразования данных*

Переменные в R. Типы данных: числовой, целочисленный, логический, текстовый. Преобразование типов. Факторы (factor vectors) и уровни. Базовые объекты в R: векторы, списки, матрицы. Векторы: создание, доступ к элементам, изменение, добавление и удаление элементов, выбор элементов, сортировка. Матрицы и списки: создание, доступ к элементам, изменение, добавление и удаление элементов.

### ***Тема 3. Основы управления данными***

Загрузка данных в R. Открытие, изменение и запись файлов. Создание базы данных. Преобразование матриц и списков в объект data.frame. Объединение баз данных. Использование команд SQL для преобразования таблиц данных

### ***Тема 4 Управляющие конструкции в R. Циклы for и while***

Операторы в R. Условные операторы if и else. Множественные и разветвленные условия. Циклы в R. Циклы for и while: достоинства и недостатки. Циклы vs векторные операции в R. Устройство функций в R. Основные функции в R. Функция assign(). Написание простейших функций в R.

### ***Тема 5. Визуализация данных***

Базовые графики в R. Функция plot(). Построение графиков математических функций. Генерирование данных в R. (Псевдо)случайные значения. Визуализация количественных данных в R: гистограммы, графики плотности распределения, ящики с усами. Визуализация качественных данных в R: таблицы сопряженности, столбчатые и круговые диаграммы. Описательные статистики. Пакет ggplot2.

### ***Тема 6. Основы корреляционно – регрессионного анализа. Кластеризация.***

Использование функций преобразования данных и отбора переменных из пакета caret. Возможности пакетов corplot и ggcorplot. Построение простейших регрессионных моделей (lm, glm, nls). Пакеты для кластерного анализа cluster, factoextra.

## **4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине**

### **4.1. Формы и методы текущего контроля успеваемости обучающихся**

В ходе реализации дисциплины «Язык программирования R» используются следующие методы текущего контроля успеваемости обучающихся:

Таблица 4.1

Тема (раздел)	Формы (методы) текущего контроля успеваемости
Тема 1. Общее представление о задачах и особенностях R.	О
Тема 2. Классы объектов, математические и статистические функции преобразования данных, ввод и вывод данных	Контрольное задание
Тема 3. Основы работы с базами данных	Контрольное задание
Тема 4. Управляющие конструкции в R. Циклы for и while. Функции в R.	Контрольное задание

Тема 5. Визуализация данных.	Расчётно-графическое задание
Тема 6. Основы корреляционно – регрессионного анализа. Кластеризация.	Расчётно-графическое задание

## Типовые материалы текущего контроля успеваемости обучающихся

### Типовые оценочные материалы по теме 1

Типовые вопросы для устного опроса по теме 1

1. Что такое "векторный подход", "векторизация"?
2. Чем отличаются матрицы и векторы как объекты R?
3. Чем отличаются матрицы и таблицы как объекты R?
4. Что такое список как объект R?
5. Извлечение данных из векторов и матриц.
6. Извлечение данных из таблиц и списков
7. Ввод данных из Excel-подобных файлов
8. Ввод данных из текстовых файлов
9. Команды вывода на экран и в текстовый файл
10. Создание скриптов.

### Типовые оценочные материалы по теме 2

Типовые контрольные задания по теме 2

1. Найти координаты вектора  $a(1, -2, 3, -4)$  в базисе  $x(1, 2, 3, 4)$ ,  $y(4, 3, 2, 1)$ ,  $z(1, 3, 4, 2)$ ,  $t(1, 4, 2, 3)$ .
2. Доказать, что система векторов  $\{(2, 1, 2), (3, -1, 4), (2, 4, 1)\}$  является базисом пространства  $R^3$  и найти координаты вектора  $x = (8, -4, 4)$  в этом базисе.
3. Даны две матрицы A и B. Найти: а) AB; б) BA; в) 3AB-2AT

$$A = \begin{pmatrix} -4 & 0 & 1 \\ 2 & -1 & 3 \\ 3 & 2 & 2 \end{pmatrix}, B = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 0 & 1 \\ -2 & 1 & 3 \end{pmatrix}.$$

4. Проверить, будет ли вектор  $a(1, 5, 0, -4)$  ортогонален базисным векторам  $x(-1, 2, 0, 4)$ ,  $y(4, -6, 2, 1)$ ,  $z(1, 3, -4, 0)$ ,  $t(7, 4, 2, -3)$ .

5. Найти ранг матрицы  $B = \begin{pmatrix} 3 & 4 & 2 \\ 1 & -1 & 1 \\ 8 & 5 & 3 \end{pmatrix}$ .

6. Найти собственные значения и собственные векторы матрицы  $A = \begin{pmatrix} 1 & 3 & 2 \\ 1 & 4 & 1 \\ 2 & 5 & 3 \end{pmatrix}$ .

7. Установить тип кривых, преобразовать уравнения кривых к каноническому виду. Определить линии и построить их.



$$ax^2 + 2bxy + cy^2 - f_1 = 0,$$

- a)  $a = 0, b = -2, c = 3, f_1 = 4$
- b)  $a = 3, b = -2, c = 0, f_1 = 4$
- c)  $a = 3, b = 2, c = 0, f_1 = 4$
- d)  $a = 0, b = 2, c = 3, f_1 = 4$
- e)  $a = 0, b = 3, c = 8, f_1 = 9$
- f)  $a = 8, b = 3, c = 0, f_1 = 9$
- g)  $a = 8, b = 3, c = 0, f_1 = 9$
- h)  $a = 0, b = 3, c = 8, f_1 = 9$
- i)  $a = 0, b = 4, c = 15, f_1 = 16$
- j)  $a = 15, b = 4, c = 0, f_1 = 16$

8. Дан вектор, в котором хранятся следующие значения:  
1, 0, 2, 3, 6, 8, 12, 15, 0, NA, NA, 9, 4, 16, 2, 0

Выведите на экран:

- первый элемент вектора;
- последний элемент вектора;
- элементы вектора с третьего по пятый включительно;
- элементы вектора, которые равны 2;
- элементы вектора, которые больше 4;
- элементы вектора, которые кратны 3 (делятся на 3 без остатка);
- элементы вектора, которые больше 4 и кратны 3;
- элементы вектора, которые или меньше 1, или больше 5;
- индексы элементов, которые равны 0;
- индексы элементов, которые не меньше 2 и не больше 8.

9. Написать код, который заменяет последний элемент вектора на пропущенное значение (NA). Ваш код должен работать для любого вектора (любой длины).

10. Написать код, который выводит на экран индексы пропущенных значений в векторе.

11. Написать код, который считает, сколько пропущенных значений в векторе.

12. Известно, что в базе данных хранятся показатели по 3 странам за 5 лет.

№	country	year
1	France	2020
2	France	2021
3	France	2022
4	France	2023
5	France	2024
6	Italy	2020
7	Italy	2021
8	Italy	2022

9	Italy	2023
10	Italy	2024
11	Spain	2020
12	Spain	2021
13	Spain	2022
14	Spain	2023
15	Spain	2024

- создайте вектор с названиями стран (первый столбец);
- создайте вектор, который мог бы послужить вторым столбцом в таблице, представленной выше (подумайте, какую длину имеет этот вектор).

### Типовые оценочные материалы по теме 3

#### Типовые контрольные задания по теме 3

Загрузите базу данных `buildings.csv` по ссылке [https://github.com/SvPolyanskaya/R\\_programming/blob/main/buildings.csv](https://github.com/SvPolyanskaya/R_programming/blob/main/buildings.csv)

1. Используя функции `dplyr`, удалите из базы данных строки, в которых значения года постройки не указаны ("н.д."), а после превратите столбец `house_age` в числовой (R будет воспринимать столбец как факторный, поэтому нужно сначала превратить его в текстовый, а затем - в числовой).
2. Посмотрите на минимальное и максимальное значения года постройки. Разумны ли они? Если нет, уберите из базы данных "странные" значения годы постройки.
3. Выберите из базы данных строки, соответствующие Басманному району, и сохраните их в базу `basn_data`.
4. Сгруппируйте данные в базе по районам и посмотрите, сколько домов в базе относятся к каждому району.
5. Найдите средний возраст дома (возраст - в годах, считаем от 2017) по каждому району. (сначала необходимо добавить переменную `house_age` - "возраст дома").
6. Найдите улицу с самым старым домом.
7. Найдите улицу с самым старым домом по каждому району и сохраните их в базу `old_house`.
8. Создайте фрейм данных из  $N = 20$  записей со следующими полями: `Nrow` – номер записи, `Name` – имя сотрудника, `Birth_Year` – год рождения, `EmployYear` – год приема на работу, `Salary` – зарплата. Заполните данный фрейм данными так, что `Nrow` изменяется от 1 до  $N$ , `Name` задается произвольно, `Birth_Year` распределен равномерно на отрезке  $[1960, 1985]$ , `Employ_Year` распределен равномерно на отрезке  $[\text{Birth\_Year} + 18, 2024]$ , `Salary` для работников младше 1990 г.р. определяется по формуле  $\text{Salary} = (\ln(2024 - \text{EmployYear}) + 1) * 28000$ , для остальных  $\text{Salary} = (\log_2(2024 - \text{EmployYear}) + 1) * 28000$ .
9. Для данных из предыдущей задачи подсчитайте число сотрудников с зарплатой, большей 75000. Добавьте в таблицу поле, соответствующее суммарному

подходящему налогу (ставка 13%), выплаченному сотрудником за время работы в организации, если его зарплата за каждый год начислялась согласно формулам для Salary, где вместо 2024 следует последовательно подставить каждый год работы сотрудника в организации.

## Типовые оценочные материалы по теме 4

### Типовые контрольные задания по теме 4

1. Написать функцию, которая просит пользователя ввести свое имя, и если оно совпадает Вашим, выводит на экран "[Name], welcome to R! I missed you", если не совпадает, то выводит на экран "Sorry, R is not available now". Подразумевается, что R знает Ваше имя (при желании можете написать функцию, которая вытаскивает Ваше имя пользователя на компьютере и сравнивает с ним).

2. Написать функцию, которая принимает на вход числовой вектор и возвращает вектор, состоящий из квадратов элементов вектора.

3. Написать функцию `smart_squares`, которая принимает на вход вектор, и если он числовой, то возвращает вектор из квадратов его элементов, а если нет -- возвращает вектор из пропущенных значений и выводит на экран сообщение "Your vector is not numeric."

4. Написать функцию, которая принимает на вход базу данных, удаляет в ней пропущенные значения и выводит на экран описательные статистики по всем переменным.

5. Написать функцию, которая принимает на вход путь к папке, в которой лежит база данных и название файла с базой данных (в csv-формате), и возвращает вектор, содержащий следующую информацию: число переменных в базе, число наблюдений в базе, количество строк, включающих пропущенные значения.

6. Напишите функцию, которая принимает на вход числовой вектор  $x$  и число разбиений интервала  $k$  (по умолчанию равное числу элементов вектора, разделенному на 10) и выполняет следующее: находит минимальное и максимальное значение элементов вектора  $x$ , разделяет полученный отрезок  $[x_{\min}; x_{\max}]$  на  $k$  равных интервалов и подсчитывает число элементов вектора, принадлежащих каждому интервалу.

## Типовые оценочные материалы по теме 5

### Типовые контрольные задания по теме 5

Загрузите базу данных `demography.csv` по ссылке [https://github.com/SvPolyanskaya/R\\_programming/blob/main/demography.csv](https://github.com/SvPolyanskaya/R_programming/blob/main/demography.csv). В ней содержатся данные по населению Белгородской и Калужской областей за 2022 год (источник - Росстат).

1. Создайте переменную `young_share` - процент населения возраста, моложе трудоспособного. Создайте переменную `trud_share` - процент населения трудоспособного возраста и `old_share` - процент населения возраста, старше трудоспособного.

2. Постройте гистограмму для доли трудоспособного населения в процентах. Измените цвет гистограммы, добавьте `rugs`. Добавьте вертикальную линию, которая очерчивает медианное значение доли трудоспособного населения в процентах.

3. Постройте гистограмму для доли населения `old_share` 3-мя разными методами (Стерджесса, Фридмана-Диакониса, Скотта).

4. Постройте сглаженные графики плотности распределения для доли трудоспособного населения в процентах по регионам (два графика в одной плоскости). Настройте цвета и прозрачность заливки. По графикам плотности определите, имеет ли смысл для визуализации распределения доли трудоспособного населения строить скрипичные диаграммы (violin plot). Если да, постройте их (так же по группам). Если нет, постройте ящики с усами.

5. Постройте диаграмму рассеяния для переменных `young_share` и `old_share`. Можно ли сказать, что чем больше процент молодого населения (моложе трудоспособного населения), тем меньше процент пожилых людей (старше трудоспособного возраста)? Поменяйте цвет и тип маркера для точек.

6. Создайте переменную `male_share`- доля мужского населения в районе/городе (в процентах). Создайте переменную `male`, которая принимает значение 1, если доля мужчин в муниципальном районе/городе больше доли женщин, и значение 0 - во всех остальных случаях.

7. Постройте пузырьковую диаграмму (bubble plot) для переменных `young_share` и `old_share`, учитывая информацию о доле мужчин в районе и о том, преобладают ли мужчины в районе или нет.

8. Постройте столбиковую диаграмму (bar plot), которая показывала бы, сколько в базе данных районов Белгородской области, а сколько -- Калужской.

## Типовые оценочные материалы по теме 6

### Типовые контрольные задания по теме 6

Загрузите файл «Успеваемость.xls» по ссылке [https://github.com/SvPolyanskaya/R\\_programming/blob/main/Успеваемость%20в%20курсах.xlsx](https://github.com/SvPolyanskaya/R_programming/blob/main/Успеваемость%20в%20курсах.xlsx). Постройте парные линейные регрессии — зависимости результативного признака  $y$  (качества) от факторов  $x_1$  и  $x_2$ , взятых по отдельности.

1. По результатам исследования оформите отчет, в котором должны быть приведены:

- a) исходные данные;
- b) матрица парных корреляций;
- c) найденное уравнение линейной регрессии;
- d) диаграмма рассеяния с линией регрессии;
- e) выводы о статистической значимости найденных коэффициентов;
- f) значение точечного прогноза для значений факторов на 200% превышающих средние выборочные для каждого из факторов.

2. Постройте множественную регрессию показателя на оба фактора вместе: перед построением множественной регрессии исследуйте мультиколлинеарность факторов.

3. Подобрать оптимальное число кластеров, разбить на кластеры и графически изобразить.

## 5. Оценочные средства для промежуточной аттестации

**Зачет с оценкой проводится с применением следующих методов (средств):**

Во время диф.зачета проверяется этап освоения компетенций ОПК 1.2; 4.1; 4.2.

Во время проверки сформированности этапа компетенции оцениваются:

- Презентация модели и полученных результатов в виде отчета или в офисных приложениях.
- Представление хода и результата решения, тестирование.
- Оценка правильности ответов на поставленные вопросы или тесты.
- Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы. Показатели и критерии оценивания компетенций с учетом этапа их формирования.

Таблица 5.1

Код компетенции	Наименование компетенции	Код компонента компетенции	Наименование компонента компетенции
ПКС 5	Способен решать задачи анализа больших данных с использованием существующих программных средств и технологий	ПКС 5.1  ПКС 5.2	Знает модели данных, основные методы получения, хранения и обработки информации  Умеет анализировать, структурировать информацию и разрабатывать базы данных информационных систем

**Показатели и критерии оценивания компетенций с учетом этапа их формирования**

Таблица 5.2

Этапы освоения компетенции	Показатели оценивания	Критерий оценивания
ПКС 5.1; 5.2	<ol style="list-style-type: none"> <li>1. Решает прикладные задачи бизнес-моделирования с использованием программирования на языке R</li> <li>2. Демонстрирует понимание используемых средств.</li> <li>3. Объясняет используемые средства, их роль в</li> </ol>	<ol style="list-style-type: none"> <li>1. Полное и правильное решение задачи.</li> <li>2. Дано объяснение полученных результатов, диапазона их использования, указаны ограничения и допущения</li> </ol>

Этапы освоения компетенции	Показатели оценивания	Критерий оценивания
	бизнес-аналитике	

Для оценки сформированности компетенций, знаний и умений, соответствующих данным компетенциям, используются контрольные вопросы, а также задачи, при решении которых необходимо написать код на языке R.

### Типовые вопросы, выносимые на зачёт с оценкой:

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. Особенные типы данных (NA, NaN, Inf).
7. Команды, связанные с отдельными типами распределений: r\*(), f\*(), d\*(), q\*().
8. Векторы, создание векторов, основные команды.
9. Матрицы, создание матриц, основные команды.
10. Массивы, создание массивов, основные команды.
11. Списки, создание списков, основные команды.
12. Факторы, создание факторов, основные команды.
13. Фреймы, создание фреймов, основные команды.
14. Ввод данных с внешних носителей.
15. Вывод данных в файлы (команды sink(), print(), cat() и другие. Их особенности).
16. Графические команды высокого уровня plot(), hist(), boxplot().
17. Графические команды низшего уровня lines(), points(), text(), legend() и другие.
18. Настройка параметров вывода графической информации.
19. Библиотека ggplot2.
20. Условный оператор if() {}, пример использования.
21. Операторы for() {}, while() {}, repeat {}, примеры использования.
22. Создание функций. Пример.
23. Подключение внешних библиотек. Примеры библиотек.
24. Работа со справочными материалами. Команды ?, help(), apropos()
25. Команды сортировки sort(), order().
26. Возможности пакетов corplot и ggcorplot.
27. Команда lm(), объект типа lm.
28. Команды кластеризации (k-means(), hclust(), cutree()...).

### Типовые задачи

1. Извлеките строки датафрейма (данные по продуктам), масса которых ниже средней массы.
2. Создайте функцию, которая вычисляет площадь круга по радиусу
3. Напишите функцию, которая будет возвращать все делители числа в виде

числового вектора

4. Выберите из датафрейма указанные колонки и отсортируйте строчки по убыванию.
5. Посчитайте количество продуктов по типам и отсортируйте по убыванию. Извлеките первые 5 строк.
6. Написать функцию, которая по уравнению  $ax^2 + 2bxy + cy^2 + dx + ey + f = 0$  выводит тип кривой, задаваемой этим уравнением.
7. Найти минимум функции  $y = x^3 - 3x^2 + 5$  на интервале  $[0; 3]$ .
8. Найти минимум функции  $z(x, y) = 100(y - x^2) - (1 + x^2)$  в области  $D = \{(x; y): x < 0,9; y - x > 0,1\}$ .
9. По данному файлу «data\_inform.csv» найти средний возраст сотрудника по каждому отделу и визуализировать, используя функцию hist.
10. По данному файлу «data\_inform.csv» найти самого молодого сотрудника по каждому отделу и визуализировать, используя функцию boxplot.
11. По данному файлу «data\_inform.csv» найти среднюю медианную зарплату сотрудника по каждому отделу и визуализировать, используя функцию barplot.
12. По данному файлу «data\_inform.csv» визуализировать данные по средней зарплате, используя функцию pie.
13. Напишите функцию, которая на основе данных из 1000 строк формирует невозвратную выборку из  $n$  элементов механическим способом.
14. Напишите функцию, которая на основе данных из 1000 строк формирует невозвратную выборку из  $n$  элементов случайным способом.
15. В указанном файле устраните данные, отклонение которых от среднего значения не превышает
16. Найти координаты вектора  $a(0, -2, 5, -4)$  в базисе  $x(1, 8, -3, 4)$ ,  $y(4, -3, 6, 1)$ ,  $z(-2, 7, 4, 2)$ ,  $t(1, 4, 2, 3)$
17. Доказать, что система векторов  $\{(2, 3, 2), (3, 1, 4), (5, 4, 1)\}$  является базисом пространства  $R^3$  и найти координаты вектора  $x = (7, 1, 4)$  в этом базисе.
18. Даны две матрицы  $A$  и  $B$ . Найти: а)  $AB$ ; б)  $BA$ ; в)  $3AB - 2AT$
19. Построить график функции  $z(x, y) = x^2 - y^2 + xy - 5$  и сохранить график в виде файла с расширением pdf.
20. Построить график функции  $z(x, y) = x^2 - y^2 + xy - 5$ , используя функцию ggplot.

### Шкала оценивания

Схема расчетов сформирована в соответствии с учебным планом направления, согласована с руководителем научно-образовательного направления, утверждена деканом факультета. Схема расчетов доводится до сведения студентов на первом занятии по данной дисциплине и является составной частью рабочей программы дисциплины и содержит информацию по изучению дисциплины, указанную в Положении о балльно-рейтинговой системе оценки знаний обучающихся в РАНХиГС.

На основании п. 14 Положения о балльно-рейтинговой системе оценки знаний обучающихся в РАНХиГС в институте принята следующая шкала перевода оценки из многобалльной системы в пятибалльную:

### **Шкала перевода оценки из многобалльной в систему «зачтено»/ «не зачтено»:**

Оценка результатов производится на основе балльно-рейтинговой системы (БРС). Использование БРС осуществляется в соответствии с приказом от 28 августа 2014 г. №168 «О применении балльно-рейтинговой системы оценки знаний студентов». БРС по дисциплине отражена в схеме расчетов рейтинговых баллов (далее – схема расчетов). Схема расчетов сформирована в соответствии с учебным планом направления, согласована с руководителем научно-образовательного направления, утверждена деканом факультета. Схема расчетов доводится до сведения студентов на первом занятии по данной дисциплине и является составной частью рабочей программы дисциплины и содержит информацию по изучению дисциплины, указанную в Положении о балльно-рейтинговой системе оценки знаний обучающихся в РАНХиГС.

На основании п. 14 Положения о балльно-рейтинговой системе оценки знаний обучающихся в РАНХиГС в институте принята следующая шкала перевода оценки из многобалльной системы в пятибалльную:

Таблица 5.3

от 0 до 50 баллов	«не зачтено»
от 51 до 100 баллов	«зачтено»

Примечание: если дисциплина изучается в течение нескольких семестров, схема расчета приводится для каждого из них.

### **6. Методические указания для обучающихся по освоению дисциплины**

Рабочей программой дисциплины предусмотрены следующие виды аудиторных занятий: лекции, практические занятия, контрольные работы. На лекциях рассматриваются наиболее сложный материал дисциплины. Лекция сопровождается презентациями, компьютерными текстами лекции, что позволяет студенту самостоятельно работать над повторением и закреплением лекционного материала. Для этого студенту должно быть предоставлено право самостоятельно работать в компьютерных классах в сети Интернет.

Практические занятия предназначены для самостоятельной работы студентов по решению конкретных задач дискретно математики. Практические занятия проводятся в компьютерных классах с использованием RStudio. Каждое практическое занятие сопровождается домашними заданиями, выдаваемыми студентам для решения внеаудиторное время. Для оказания помощи в решении задач имеются тексты практических заданий с условиями задач и вариантами их решения.

С целью контроля сформированности компетенций разработан фонд контрольных заданий. Его использование позволяет реализовать балльно-рейтинговую оценку, определенную приказом от 28 августа 2014 г. №168 «О применении балльно-рейтинговой системы оценки знаний студентов».

Для активизации работы студентов во время контактной работы с преподавателем отдельные занятия проводятся в интерактивной форме. В основном интерактивная форма занятий обеспечивается при проведении занятий в компьютерном классе. Интерактивная форма обеспечивается наличием разработанных заданий, наличием контрольных вопросов, возможностью доступа к системе дистанционного обучения, а также к тестеру.

Для работы с печатными и электронными ресурсами СЗИУ имеется возможность доступа к электронным ресурсам. Организация работы студентов с электронной библиотекой указана на сайте института (странице сайта – «Научная библиотека»).



## **7. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине**

### **7.1. Основная литература**

1. Кабаков Р. И.: R в действии. Анализ и визуализация данных на языке R М.: ДМК Пресс, 2014. – 588 с. (ISBN 978-5-947060-077-1)
2. Мэтлофф Н. Искусство программирования на R. Погружение в большие данные / Мэтлофф Норман – СПб.: Питер, 2019. – 416 с.
3. Шипунов А.Б., Балдин Е.М., Волкова м и др.: Наглядная статистика. Используем R! ДМК Пресс, 2014, 298 с. (ISBN: 978-5-97060-094-8)

### **7.2. Дополнительная литература**

1. Гришин В. А., Тихов М. С. Методы обработки данных и моделирование на языке R: Учебно-методическое пособие / Гришин В.А., Тихов М.С. — Нижний Новгород: Нижегородский госуниверситет, 2019. – 54с.
2. Джеймс Г., Уиттон Д., Хасты Т., Тибширани Р.: Введение в статистическое обучение с примерами на языке R. (ISBN: 978-5-97060-293-5 )
3. Дуглас Люк: Анализ сетей (графов) в среде R. Руководство пользователя. (перевод с английского, ISBN: 978-5-97060-428-1 )
4. Мастицкий С.Э., Шитиков В.К.: Статистический анализ и визуализация данных с помощью R (ISBN: 978-5-97060-301-7)
5. Шипунов А.Б. R - объектно-ориентированная статистическая среда, <http://herba.msu.ru/shipunov/software/r/r-ru.htm>
6. Меретиллов М.А. Методические указания к лабораторным работам по курсу «Методы анализа данных», <http://gislab.info/docs/r-metoda-2006.10.23.pdf>

### **7.3. Учебно-методическое обеспечение самостоятельной работы.**

1. Положение об организации самостоятельной работы студентов федерального государственного бюджетного образовательного учреждения высшего образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации» (в ред. приказа РАНХиГС от 11.05.2016 г. № 01–2211);
2. Положение о курсовой работе (проекте), выполняемой студентами федерального государственного бюджетного образовательного учреждения высшего образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации» (в ред. приказа РАНХиГС от 11.05.2016 г. № 01–2211).

### **7.4. Нормативные правовые документы**

Не используются

### **7.5. Интернет-ресурсы**

СЗИУ располагает доступом через сайт научной библиотеки <http://nwapa.spb.ru/> к следующим подписным электронным ресурсам:

## Русскоязычные ресурсы

Электронные учебники электронно - библиотечной системы (ЭБС) «Айбукс»

Электронные учебники электронно – библиотечной системы (ЭБС) «Лань»

Рекомендуется использовать следующий интернет-ресурсы

<http://serg.fedosin.ru/ts.htm>

<http://window.edu.ru/resource/188/64188/files/chernyshov.pdf>

### 7.6. Иные источники

Не используются.

## **8. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы**

Все практические занятия проводятся в компьютерном классе. Учебная дисциплина включает использование программного обеспечения Microsoft Excel, Microsoft Word, Microsoft Power Point для подготовки текстового и табличного материала, графических иллюстраций.

Методы обучения с использованием информационных технологий (компьютерное тестирование, демонстрация мультимедийных материалов).

Интернет-сервисы и электронные ресурсы (поисковые системы, электронная почта, профессиональные тематические чаты и форумы, системы аудио и видео конференций, онлайн энциклопедии, справочники, библиотеки, электронные учебные и учебно-методические материалы).

Система дистанционного обучения Moodle.