

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

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

Экономический факультет  
Кафедра экономической кибернетики и прикладной статистики

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



**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**

**«МНОГОМЕРНЫЙ АНАЛИЗ И ПРОГНОЗИРОВАНИЕ СРЕДСТВАМИ  
ЯЗЫКА ПРОГРАММИРОВАНИЯ R»**

По направлению подготовки 38.04.05 Бизнес-информатика  
Программа магистратуры: «Экономическая аналитика и бизнес-статистика»

## Лист согласования РПУД

Рабочая программа учебной дисциплины «Многомерный анализ и прогнозирование средствами языка программирования R» по направлению подготовки 38.04.05 Бизнес-информатика. – 30 с.

Рабочая программа учебной дисциплины «Многомерный анализ и прогнозирование средствами языка программирования R» составлена с учетом Федерального государственного образовательного стандарта высшего образования по направлению подготовки 38.04.05 Бизнес-информатика Министерства науки и высшего образования Российской Федерации от 12 августа 2020 г. № 990.

СОСТАВИТЕЛЬ (СОСТАВИТЕЛИ):

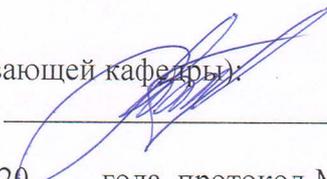
к.э.н., доцент Спорняк С.А.

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

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

Переутверждена: «  »    20   г., протокол №   

Согласована (для обеспечивающей кафедры):

Декан экономического факультета  Тхор Е.С.

Переутверждена: «  »    20   года, протокол №   

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

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

## Структура и содержание дисциплины

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

Основной целью дисциплины «Многомерный анализ и прогнозирование средствами языка программирования R» является формирование у студентов углубленных знаний в области математических методов анализа экономики; применения современных методов анализа данных для получения прогнозов; применение современных методов анализа данных в прикладных задачах.

Исходя из цели, в процессе изучения дисциплины решаются следующие задачи:

познакомить с основными алгоритмами многомерного статистического анализа;

обсудить примеры применения этих алгоритмов как в экономических исследованиях, так и в практических задачах с помощью языка программирования R.

### 2. Место дисциплины в структуре ОПОП ВО

Дисциплина «Многомерный анализ и прогнозирование средствами языка программирования R» относится к вариативной части (часть, формируемая участниками образовательных отношений) дисциплин.

Необходимыми условиями для освоения дисциплины являются: знания основных принципов организации информационных систем, теории вероятностей и математической статистики, навыки обработки статистических данных с помощью языка программирования R.

Содержание дисциплины является логическим продолжением содержания дисциплин «Бизнес-информатика», «Бизнес-информатика 2», «Статистика», «Эконометрика», «Прогнозирование социально-экономических процессов», «Современные технологии обработки статистических данных», изучаемых в рамках бакалавриата, и служит основой для освоения дисциплин: «Математические методы и модели рыночной экономики», написания магистерской диссертации.

### 3. Требования к результатам освоения содержания дисциплины

Код и наименование компетенции	Индикаторы достижений компетенции (по реализуемой дисциплине)	Перечень планируемых результатов
ПК-2. Способен применять экономико-математические методы для моделирования социально-экономических систем	ПК-2.2. Способен проводить анализ социально-экономических систем с применением языков высокого уровня	<b>Знать:</b> методы анализа социально-экономических систем с применением языков высокого уровня
		<b>Уметь:</b> проводить анализ социально-экономических систем с применением языков высокого уровня
		<b>Владеть:</b> навыками проведения анализа социально-экономических систем с применением языков высокого уровня

## 4. Структура и содержание дисциплины

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

Вид учебной работы	Очная форма	Заочная форма
Общая учебная нагрузка (всего)	144 (4 зач. ед)	144 (4 зач. ед)
Обязательная контактная работа (всего) в том числе:	56	20
Лекции	28	12
Семинарские занятия	-	-
Практические занятия	28	8
Лабораторные работы	-	
Курсовая работа (курсовой проект)		
Другие формы и методы организации образовательного процесса ( <i>расчетно-графические работы, индивидуальные задания и т.п.</i> )	-	-
Самостоятельная работа студента (всего)	88	124
Форма аттестации	экзамен	экзамен

### 4.2. Содержание разделов дисциплины

#### Тема 1. ПРЕДМЕТ И ЗАДАЧИ КУРСА

Предмет и задачи дисциплины «Многомерный анализ и прогнозирование средствами языка программирования R», строение дисциплины, организация изучения, основная и дополнительная литература, практика использования многомерных методов в статистическом анализе. Состав учебно-методического обеспечения дисциплины. Реализация моделей Data Mining в среде R.

#### Тема 2. СТАТИСТИЧЕСКИЕ МОДЕЛИ: КРИТЕРИИ И МЕТОДЫ ИХ ОЦЕНИВАНИЯ

Основные шаги построения и верификации моделей. Использование алгоритмов ресэмплинга для тестирования и оптимизации параметров моделей. Модели для предсказания класса объектов. Проецирование многомерных данных на плоскости. Многомерный статистический анализ данных. Методы кластеризации.

#### Тема 3. ПАКЕТ CARET - ИНСТРУМЕНТ ПОСТРОЕНИЯ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ В R

Универсальный интерфейс доступа к функциям машинного обучения в пакете caret. Обнаружение и удаление "ненужных" предикторов. Предварительная обработка данных: преобразование и групповая трансформация переменных. Заполнение пропущенных значений в данных. Функция train() пакета caret.

#### Тема 4. ПОСТРОЕНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ РАЗЛИЧНОГО ТИПА

Селекция оптимального набора предикторов линейной модели. Регуляризация, частные наименьшие квадраты и kNN-регрессия. Построение

деревьев регрессии. Ансамбли моделей: бэггинг, случайный лес, бустинг. Сравнение построенных моделей и оценка информативности предикторов Деревья регрессии с многомерным откликом.

#### **Тема 5. БИНАРНЫЕ МАТРИЦЫ И АССОЦИАТИВНЫЕ ПРАВИЛА**

Классификация в бинарных пространствах с использованием классических моделей. Бинарные деревья решений. Поиск логических закономерностей в данных. Алгоритмы выделения ассоциативных правил. Анализ последовательностей знаков или событий.

#### **Тема 6. БИНАРНЫЕ КЛАССИФИКАТОРЫ С РАЗЛИЧНЫМИ РАЗДЕЛЯЮЩИМИ ПОВЕРХНОСТЯМИ**

Дискриминантный анализ. Метод опорных векторов. Ядерные функции машины опорных векторов. Деревья классификации, случайный лес и логистическая регрессия. Процедуры сравнения эффективности моделей классификации.

#### **Тема 7. МОДЕЛИ КЛАССИФИКАЦИИ ДЛЯ НЕСКОЛЬКИХ КЛАССОВ**

Ирисы Фишера и метод k-ближайших соседей. Наивный классификатор Байеса. Классификация в линейном дискриминантном пространстве. Нелинейные классификаторы в R. Модель мультиномиального логита. Классификаторы на основе искусственных нейронных сетей.

#### **Тема 8. МОДЕЛИРОВАНИЕ ПОРЯДКОВЫХ И СЧЕТНЫХ ПЕРЕМЕННЫХ**

Модель логита для порядковой переменной. Настройка параметров нейронных сетей средствами пакета caret. Методы комплексации модельных прогнозов. Обобщенные линейные модели для счетных данных. ZIP- и барьерные модели счетных данных

#### **Тема 9. МЕТОДЫ МНОГОМЕРНОЙ ОРДИНАЦИИ**

Преобразование данных и вычисление матрицы расстояний. Непараметрический дисперсионный анализ матриц дистанций. Методы ординации объектов и переменных: построение и сравнение диаграмм. Оценка связи ординации с внешними факторами. Неметрическое многомерное шкалирование и построение распределения чувствительности видов.

#### **Тема 10. КЛАСТЕРНЫЙ АНАЛИЗ**

Алгоритмы кластеризации, основанные на разделении. Иерархическая кластеризация. Оценка качества кластеризации. Другие алгоритмы кластеризации. Самоорганизующиеся карты Кохонена.

#### **Тема 11. RATTLE: ГРАФИЧЕСКИЙ ИНТЕРФЕЙС R ДЛЯ РЕАЛИЗАЦИИ АЛГОРИТМОВ DATA MINING**

Начало работы с пакетом rattle. Описательная статистика и визуализация данных. Построение и тестирование моделей классификации. Дескриптивные модели (обучение без учителя).

### 4.3. Лекции

№ п/п	Название темы	Очная форма	Заочная форма
1.	Предмет и задачи курса	2	1
2.	Статистические модели: критерии и методы их оценивания	3	1
3.	Пакет caret - инструмент построения статистических моделей в R	3	2
4.	Построение регрессионных моделей различного типа	3	1
5.	Бинарные матрицы и ассоциативные правила	3	1
6.	Бинарные классификаторы с различными разделяющими поверхностями	3	1
7.	Модели классификации для нескольких классов	3	1
8.	Моделирование порядковых и счетных переменных	2	1
9.	Методы многомерной ординации	2	1
10.	Кластерный анализ	2	1
11.	Rattle: графический интерфейс R для реализации алгоритмов Data Mining	2	1
<b>Итого:</b>		<b>28</b>	<b>12</b>

### 4.4. Практические (семинарские) занятия

№ п/п	Название темы	Очная форма	Заочная форма
1.	Предмет и задачи курса	2	0,5
2.	Статистические модели: критерии и методы их оценивания	3	0,5
3.	Пакет caret - инструмент построения статистических моделей в R	3	0,5
4.	Построение регрессионных моделей различного типа	3	0,5
5.	Бинарные матрицы и ассоциативные правила	3	0,5
6.	Бинарные классификаторы с различными разделяющими поверхностями	3	0,5
7.	Модели классификации для нескольких классов	3	1
8.	Моделирование порядковых и счетных переменных	2	1
9.	Методы многомерной ординации	2	1
10.	Кластерный анализ	2	1
11.	Rattle: графический интерфейс R для реализации алгоритмов Data Mining	2	1
<b>Итого:</b>		<b>28</b>	<b>8</b>

### 4.5. Лабораторные работы

Планом не предусмотрены.

#### 4.6. Самостоятельная работа студентов

№ п/п	Название темы	Вид СРС	Очная форма	Заочная форма
1.	Предмет и задачи курса	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	4	8
2.	Статистические модели: критерии и методы их оценивания	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	6	8
3.	Пакет caret – инструмент построения статистических моделей в R	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	4	8
4.	Построение регрессионных моделей различного типа	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	4	8
5.	Бинарные матрицы и ассоциативные правила	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	4	8
6.	Бинарные классификаторы с различными разделяющими поверхностями	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	4	8
7.	Модели классификации для нескольких классов	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	6	8
8.	Моделирование порядковых и счетных переменных	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	4	8
9.	Методы многомерной ординации	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	6	8
10.	Кластерный анализ	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	6	8
11.	Rattle: графический интерфейс R для реализации алгоритмов Data Mining	Подготовка к практическим занятиям, к текущему и промежуточному контролю знаний	4	8
12.	Экзамен	Подготовка к экзамену	36	36
<b>Итого:</b>			<b>88</b>	<b>124</b>

**4.7. Курсовые работы/проекты не предполагаются учебным планом.**

#### 5. Образовательные технологии

Преподавание дисциплины ведется с применением следующих видов образовательных технологий: объяснительно-иллюстративного обучения

(технология поддерживающего обучения, технология проведения учебной дискуссии), информационных технологий (презентационные материалы), развивающих и инновационных образовательных технологий.

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

## 6. Формы контроля освоения дисциплины

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

- вопросы для обсуждения (в виде докладов и сообщений);
- контрольные работы;
- тесты.

Промежуточная аттестации по результатам освоения дисциплины проходит в форме экзамена (включает в себя ответы на теоретические вопросы). Студенты, выполнившие 75% текущих и контрольных мероприятий на «отлично», а остальные 25 % на «хорошо», имеют право на получение итоговой оценки.

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

Шкала оценивания (зачет с оценкой)	Характеристика знания предмета и ответов
отлично (5)	Студент глубоко и в полном объеме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач.
хорошо (4)	Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.
удовлетворительно (3)	Студент знает только основной программный материал, допускает неточности, недостаточно четкие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.
неудовлетворительно (2)	Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.

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

### **а) основная литература:**

1. Шитиков В.К., Мастицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. – 2017. – 351 с. – URL: <https://github.com/ranalytics/data-mining>
2. Мельниченко А.С. Математическая статистика и анализ данных: учеб. пособие / А.С. Мельниченко - М.: МИСиС, 2018. - 45 с. - ISBN 978-5-906953-62-9 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785906953629.html>
3. Афонин П.Н. Статистический анализ с применением современных программных средств: учебное пособие / Афонин П.Н., Афонин Д.Н. - СПб.: ИЦ Интермедия, 2017. - 100 с. - ISBN 978-4383-0080-9 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN978438300809.html>
4. Буре В.М., Парилина Е.М., Седаков А.А. Методы прикладной статистики в R и Excel: учебное пособие, Изд. Европейского университета в СПб., 2005.

### **б) дополнительная литература:**

1. Крупин В.Г. Высшая математика. Теория вероятностей, математическая статистика, случайные процессы. Сборник задач с решениями: учебное пособие / Крупин В.Г. - М.: Издательский дом МЭИ, 2019. - ISBN 978-5-383-01225-3 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785383012253.html>
2. Болдырева Н.П. Статистика в схемах и таблицах. Ч. 2 / Болдырева Н.П. - М.: ФЛИНТА, 2019. - 134 с. - ISBN 978-5-9765-1936-7 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785976519367.html>
3. Васина М.В. Теория вероятностей и математическая статистика. Руководство по решению задач. Ч. 1 / Васина М.В., Васин А.А., Манохин Е.В. - М.: Прометей, 2018. - 160 с. - ISBN 978-5-907003-70-5 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785907003705.html>
4. Завьялов О.Г. Теория вероятностей и математическая статистика с применением Excel и Maxima: Учебное пособие / Завьялов О.Г., Подповетная Ю.В. - М.: Прометей, 2018. - 290 с. - ISBN 978-5-907003-44-6 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785907003446.html>
5. Мелехина Т.Л. Лекции по теории вероятностей и математической статистике (для слушателей Института сокращенных программ): учебное пособие для бакалавров направлений 38.03.01 "Экономика", 38.03.02 "Менеджмент", 38.03.04 "Государственное и муниципальное управление" / Т.Л. Мелехина - М.: Прометей, 2018. - 130 с. - ISBN 978-5-907003-63-7 -

Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785907003637.html>

6. Тагиева Р.Ф. Обработка экспериментальных данных. Ч.1: учебное пособие: в 2 ч. / Р.Ф. Тагиева, А.Н. Титов - Казань: Издательство КНИТУ, 2017. - 96 с. - ISBN 978-5-7882-2261-5 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785788222615.html>

7. Александровская Ю.П. Многомерный статистический анализ в экономике: учебное пособие / Александровская Ю. П. - Казань: Издательство КНИТУ, 2017. - 96 с. - ISBN 978-5-7882-2191-5 - Текст: электронный // ЭБС "Консультант студента": [сайт]. - URL: <http://www.studentlibrary.ru/book/ISBN9785788221915.html>

#### **в) методические рекомендации:**

##### **г) Интернет-ресурсы:**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

15. Руководство по своду знаний по бизнес-анализу (ВАВОК 2.0.) (на рус.яз.). URL: <http://iiba.ru/chapter-1-introduction/>

16. Системы дистанционного обучения кафедры экономической кибернетики и прикладной статистики Луганского национального университета имени Владимира Даля в среде Moodle. URL: <https://ecpsdahl/ru/>

### **Информационный ресурс библиотеки образовательной организации**

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

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

Для проведения лекционных занятий требуется аудитория на курс, оборудованная мультимедийным проектором с экраном.

Для проведения лабораторных занятий требуется компьютерный класс, подключенный к Интернет.

Программное обеспечение:

<b>Функциональное назначение</b>	<b>Бесплатное программное обеспечение</b>	<b>Ссылки</b>
Офисный пакет	OpenOffice 4.3.7	<a href="https://www.openoffice.org/">https://www.openoffice.org/</a>
Операционная система	UBUNTU 19.04	<a href="https://ubuntu.com/">https://ubuntu.com/</a>
Браузер	Firefox Mozilla	<a href="http://www.mozilla.org/ru/firefox/fx">http://www.mozilla.org/ru/firefox/fx</a>
Графический редактор	GIMP (GNU Image Manipulation Program)	<a href="http://www.gimp.org/">http://www.gimp.org/</a> <a href="http://gimp.ru/viewpage.php?page_id=8">http://gimp.ru/viewpage.php?page_id=8</a> <a href="http://ru.wikipedia.org/wiki/GIMP">http://ru.wikipedia.org/wiki/GIMP</a>
Редактор PDF	Adobe Acrobat Reader	<a href="https://get.adobe.com/ru/reader/">https://get.adobe.com/ru/reader/</a>
Аудиоплеер	VLC	<a href="http://www.videolan.org/vlc/">http://www.videolan.org/vlc/</a>

### **9. Оценочные средства по дисциплине**

#### **Паспорт**

#### **оценочных средств по учебной дисциплине**

«Многомерный анализ и прогнозирование средствами языка программирования R»

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

<b>№ п/п</b>	<b>Код контролируемой компетенции</b>	<b>Формулировка контролируемой компетенции</b>	<b>Индикаторы достижений компетенции (по реализуемой дисциплине)</b>	<b>Контролируемые темы учебной дисциплины, практики</b>	<b>Этапы формирования (семестр изучения)</b>
1	ПК-2	Способен проводить анализ социально-экономических систем с применением	ПК-2.2	Тема 1	1
				Тема 2	
				Тема 3	
				Тема 4	
				Тема 5	
				Тема 6	

		языков высокого уровня		Тема 7	
				Тема 8	
				Тема 9	
				Тема 10	
				Тема 11	

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

№ п/п	Код контролируемой компетенции	Индикаторы достижений компетенции (по реализуемой дисциплине)	Перечень планируемых результатов	Контролируемые темы учебной дисциплины	Наименование оценочного средства
1.	ПК-2	ПК-2.2	Знать: методы анализа социально-экономических систем с применением языков высокого уровня Уметь: проводить анализ социально-экономических систем с применением языков высокого уровня Владеть: навыками проведения анализа социально-экономических систем с применением языков высокого уровня	Тема 1-11	Вопросы для обсуждения (в виде докладов и сообщений), контрольные работы, тесты

**Фонды оценочных средств по дисциплине «Многомерный анализ и прогнозирование средствами языка программирования R»**

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

1. Многомерный статистический анализ: основные понятия и методы.
2. Практическое использование многомерных методов в статистическом анализе.
3. Реализация моделей Data Mining в среде R.
4. Основные шаги построения и верификации моделей.

5. Использование алгоритмов ресэмплинга для тестирования и оптимизации параметров моделей.
6. Модели для предсказания класса объектов.
7. Проецирование многомерных данных на плоскости.
8. Многомерный статистический анализ данных.
9. Анализ зависимостей в многомерных данных. Регрессионный анализ.
10. Методы кластеризации.
11. Универсальный интерфейс доступа к функциям машинного обучения в пакете caret.
12. Обнаружение и удаление "ненужных" предикторов.
13. Предварительная обработка данных: преобразование и групповая трансформация переменных.
14. Заполнение пропущенных значений в данных.
15. Функция train() пакета caret.
16. Селекция оптимального набора предикторов линейной модели.
17. Регуляризация, частные наименьшие квадраты и kNN-регрессия.
18. Построение деревьев регрессии.
19. Ансамбли моделей: бэггинг, случайный лес, бустинг.
20. Сравнение построенных моделей и оценка информативности предикторов
21. Деревья регрессии с многомерным откликом.
22. Классификация в бинарных пространствах с использованием классических моделей.
23. Бинарные деревья решений.
24. Поиск логических закономерностей в данных.
25. Алгоритмы выделения ассоциативных правил.
26. Анализ последовательностей знаков или событий.
27. Дискриминантный анализ.
28. Метод опорных векторов.
29. Ядерные функции машины опорных векторов.
30. Деревья классификации, случайный лес и логистическая регрессия.  
Процедуры сравнения эффективности моделей классификации.

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

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Доклад (сообщение) представлен(о) на высоком уровне (студент в полном объеме осветил рассматриваемую проблематику, привел аргументы в пользу своих суждений, владеет профильным понятийным (категориальным) аппаратом и т.п.)
4	Доклад (сообщение) представлен(о) на среднем уровне (студент в целом осветил рассматриваемую проблематику, привел аргументы в пользу своих суждений, допустив некоторые неточности и т.п.)
3	Доклад (сообщение) представлен(о) на низком уровне (студент

	допустил существенные неточности, изложил материал с ошибками, не владеет в достаточной степени профильным категориальным аппаратом и т.п.)
2	Доклад (сообщение) представлен(о) на неудовлетворительном уровне или не представлен (студент не готов, не выполнил задание и т.п.)

### **Контрольные работы (по вариантам)**

Решение практических задач в рамках выполнения контрольной работы осуществляется с целью проверки уровня знаний, умений, владений, понимания студентом основных методов и законов изучаемой теории при решении конкретных практических задач, умения применять на практике полученных знаний.

Контрольная работа включает выполнение заданий по вариантам и содержит тематические разделы согласно перечню практических работ рабочей программы дисциплины.

Студент сдает выполненную контрольную работу (комплект заданий по темам и вариантам) с защитой.

При выполнении контрольной работы разрешено пользоваться курсом лекций или учебниками.

Задания должны быть сданы до зачетной недели.

### **Пример задач**

Задание 1.

В файле - различные характеристики банков. Требуется сделать классификацию

- с помощью иерархической процедуры
- с помощью метода k-средних

В первом случае в отчете поместить дерево, выбранное вами разумное разбиение на кластеры, указать выбор расстояния.

Во втором случае указать полученное разбиение на группы, результаты дисперсионного анализа (и прокомментировать их).

Задание 2.

В файле - различные характеристики авиакомпаний. Требуется сделать классификацию - с помощью иерархической процедуры - с помощью метода k-средних По сделанной классификации осуществить дискриминантный анализ. Привести критерии качества разбиения. Классифицировать 1-ую авиакомпанию.

Задание 3. Студентам выдается файл с характеристиками стран. Требуется выделить главные компоненты, интерпретировать их, произвести факторный анализ.

Задание 4. Студентам выдается файл с данными, необходимо произвести корреляционный, кластерный и компонентный анализ, интерпретировать результаты.

**Критерии и шкала оценивания по оценочному средству «контрольные работы (по вариантам)»**

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Задание выполнено на высоком уровне (правильные ответы даны на 90-100% вопросов/задач)
4	Задание выполнено на среднем уровне (правильные ответы даны на 75-89% вопросов/задач)
3	Задание выполнено на низком уровне (правильные ответы даны на 50-74% вопросов/задач)
2	Задание выполнено на неудовлетворительном уровне (правильные ответы даны менее чем на 50%)

**Тесты**

1. В каком методе прикладного статистического анализа используются собственные векторы и собственные значения корреляционной матрицы? Выберите правильный ответ:
  - 1) дискриминантный анализ;
  - 2) кластерный анализ;
  - 3) корреляционный анализ;
  - 4) факторный анализ.
  
2. В качестве оценки математического ожидания многомерной случайной величины принимают:
  - 1) собственный вектор;
  - 2) вектор средних;
  - 3) среднее значение;
  - 4) нет верного ответа.
  
3. Что понимается под вектором средних?
  - 1) вектор математических ожиданий;
  - 2) среднеарифметический вектор;
  - 3) среднее значение реализаций многомерных случайных величин;
  - 4) нет верного ответа.
  
4. Какой критерий используется для сравнения средних двух многомерных выборочных совокупностей?
  
5. Какие статистика и критерий используются для сравнения дисперсий двух многомерных выборочных совокупностей?
  
6. В каком методе прикладного статистического анализа используется критерий Хоттелинга сравнения средних двух многомерных выборочных совокупностей? Выберите правильный ответ:
  - а) дискриминантный анализ;

- б) кластерный анализ;
- в) корреляционный анализ;
- г) факторный анализ.

7. В каком методе прикладного статистического анализа используется матрица близостей (расстояний) типа «объект - объект»?

8. Тест на симметричность распределения в факторном анализе применяется в случае наличия

- 1) засорения
- 2) линейной зависимости
- 3) автокорреляции
- 4) линейной независимости

9. Редуцированная корреляционная матрица, используемая в факторном анализе, является

- 1) симметрической
- 2) квадратной
- 3) вырожденной
- 4) отрицательно определенной

10. Критерий качества прогноза  $S/\sigma$  получился равным 0,4. Является ли прогноз:

- а) хорошим;
- б) удовлетворительным;
- в) неудовлетворительным.

11. Выявление признаков, влияние которых в будущем будет снижаться или возрастать в факторном анализе производится с помощью

- 1) динамических моделей
- 2) главных факторов
- 3) главных компонент
- 4) вероятностных зависимостей

12. Матрица факторного отображения, включающая нагрузки характерных факторов, называется

- 1) полной факторной матрицей
- 2) определяющей факторной матрицей
- 3) матрицей корреляции факторов
- 4) подобной матрицей

13. Сумма относительных вкладов всех общих факторов в дисперсию рассматриваемого признака у называется

- 1) общностью признака у
- 2) удельным весом признака у

- 3) надежностью признака  $u$
- 4) отклонением признака  $u$

14. Подход, при котором вначале определяется общность, а потом число факторов, используется в методе

- 1) центроидном
- 2) главных компонент
- 3) факторного обобщения
- 4) факторного сближения

15. Количество нулей в каждом столбце матрицы факторной структуры должно быть не менее

- 1) числа общих факторов
- 2) ранга матрицы факторной структуры
- 3) числа линейно независимых строк матрицы факторной структуры
- 4) числа линейно независимых столбцов матрицы факторной структуры

16. Чему равна обобщенная дисперсия вектора главных компонент?

17. Обобщенная дисперсия и сумма дисперсий компонент в результате ортогонального преобразования случайного центрированного вектора

- 1) не меняются
- 2) увеличиваются
- 3) уменьшаются
- 4) либо увеличиваются, либо уменьшаются

18. Собственные векторы, соответствующие попарно различным характеристическим числам, всегда

- 1) линейно независимы
- 2) линейно зависимы
- 3) пропорциональны
- 4) обратно пропорциональны

19. Каким методом производится определение коэффициентов характеристического многочлена по следам степеней матрицы в методе главных компонент?

20. Каким методом производится одновременное определение скалярных коэффициентов характеристического многочлена и матричных коэффициентов присоединенной матрицы в методе главных компонент?

21. Показатель факторизации в методе главных компонент (выбрать нужное)

- 1) прямо пропорционален полному вкладу всех извлеченных главных компонент в общую дисперсию всех признаков

- 2) обратно пропорционален общему количеству главных компонент
- 3) обратно пропорционален полному вкладу всех извлеченных главных компонент в общую дисперсию всех признаков
- 4) прямо пропорционален общему количеству главных компонент

22. В общем алгоритме факторного анализа укажите правильную последовательность шагов этого алгоритма:

- 1) получение редуцированной матрицы ковариаций или парных корреляций ( $R_h$ );
- 2) получение матрицы стандартизованных значений признаков ( $Z$ );
- 3) вычисление матрицы ковариаций или матрицы корреляций ( $S, R$ );
- 4) вычисление матрицы факторного отображения ( $A$ );
- 5) формирование матрицы исходных данных ( $X$ );
- 6) вычисление матрицы значений факторов ( $F$ );
- 7) получение матрицы факторного отображения после поворота ( $W$ ).

23. Верно ли следующее утверждение:

«В ходе вычислительных процедур Метода главных компонент получают все главные компоненты и их число первоначально равно числу элементарных признаков»

- 1) верно;
- 2) не верно

24. Назовите основные проблемы факторного анализа.

25. Назовите отличительные черты иерархического кластерного анализа.

26. В алгоритме иерархического кластерного анализа укажите правильную последовательность шагов этого алгоритма:

- 1) рассчитывается матрица расстояний или матрица мер сходства;
- 2) находится пара самых близких кластеров. По выбранному алгоритму объединяются эти два кластера. Новому кластеру присваивается меньший из номеров объединяемых кластеров;
- 3) значения исходных переменных нормируются одним из способов;
- 4) процедуры 2, 3 и 4 повторяются до тех пор, пока все объекты не будут объединены в один кластер или до достижения заданного «порога» сходства.

27. Какими методами определяется мера сходства для объединения двух кластеров в иерархических методах?

28. Какие методы относятся к алгоритмам классификации иерархического кластерного анализа?

29. Что является основной задачей дискриминантного анализа?

30. Что можно отнести к основным предпосылкам дискриминантного анализа?

31. Верно ли следующее утверждение:

«Стандартизованные коэффициенты в дискриминантном анализе применяют в тех случаях, когда необходимо определить, какая из используемых переменных вносит наибольший вклад в величину дискриминантной функции».

- 1) верно;
- 2) не верно.

32. Уменьшение размерности приводит

- 1) к потере данных;
- 2) к изменению данных;
- 3) к измерению данных;
- 4) к расслоению данных.

33. Анализ главных компонент – самый известный метод уменьшения \_\_\_\_\_ . (размерности)

34. Эта функция умеет самостоятельно поводить z-преобразования, для чего нужно поставить center = TRUE и scale. = TRUE. (prcomp())

35. Функция \_\_\_\_\_, примененная на результат проведения АГК, выдаст информацию о полученных компонентах. (summary()).

36. В R бутстреп легко реализуется с помощью функции \_\_\_\_\_, генерирующей любые случайные выборки с "возвращением".

Вероятность того, что конкретное наблюдение не войдет в бутстреп-выборку размера n, равна  $(1 - 1/n)^n$  и стремится к  $1/e = 0.368$  при  $n \rightarrow \infty$ :

```
Nboot = 1000; n=100  
mean(replicate(Nboot, length(unique(sample(n, replace=TR))))))
```

(sample(..., replace=T)).

37. \_\_\_\_\_ классификатора может также оцениваться с использованием информационных критериев – энтропии  $E = -\sum p_i \log_2 p_i$ , где  $p_i$  – вероятности каждого возможного исхода, и условной энтропии (conditional entropy).

**(Эффективность).**

38. Общепринятым графоаналитическим методом оценки качества теста и интерпретации перечисленных показателей является ROC-анализ. В случае идеального классификатора ROC-кривая проходит вблизи верхнего левого угла, где доля истинно-положительных случаев равна 1, а доля ложно-

положительных примеров равна нулю. А близость ROC-кривой к диагонали говорит о низкой эффективности построенной модели.

**(Верно / Неверно).**

39. Специальная функция `findLinearCombos()` предназначена для:

- 1) нахождения и исключения переменных, связанных линейными зависимостями;
- 2) разбиения исходных данных на обучающую и контрольную выборки;
- 3) для визуализации данных;
- 4) все ответы верны.

40. Большинство наиболее распространенных способов преобразования количественных предикторов может быть реализовано функцией \_\_\_\_\_ из пакета `caret`. Ее основными аргументами являются следующие: `preProcess(x, method = c("center", "scale"), na.remove = TRUE, verbose = FALSE)`.

**( `preProcess()` )**

41. Для приведения всех переменных к одинаковым единицам измерения Служит \_\_\_\_\_, являющаяся самой распространенной операцией предобработки (и потому задается в `preProcess()` по умолчанию). Она заключается в вычитании из исходного значения  $x_i$  некоторой переменной  $X$  соответствующего среднего значения  $\bar{x}$  ("центрирование", или "center") и последующего деления полученной разницы на стандартное отклонение этой переменной  $\sigma_x$  ("нормализация", или "scale"):  $x'_i = (x_i - \bar{x}) / \sigma_x$ .

**(стандартизация)**

42. Для реализации PCA-процедуры с помощью `preProcess()` необходимо определить параметр `thresh` – кумулятивную долю дисперсии исходных данных, которая должна содержаться в главных компонентах (по умолчанию `thresh = 0.95`) или `pcaComp` – максимальное число главных компонент (по умолчанию этот параметр имеет значение `NULL`, т.е. он выключен, и оптимальное число главных компонент выбирается на основе `thresh`).

**(Верно / Неверно)**

43. Оптимальная размерность пространства главных компонент определяется по критерию Кайзера-Гуттмана, который рекомендует оставить только те главные компоненты, собственные значения которых превышают их

- 1) среднее;
- 2) моду;
- 3) медиану;
- 4) нет верного ответа.

44. Вызов какой функции со значениями, принимаемыми по умолчанию, имеет следующий вид:

```

trainControl(method = "boot",
  number = ifelse(grepl("cv", method), 10, 25), p = 0.75,
  repeats = ifelse(grepl("cv", method), 1, number),
  search = "grid", initialwindow = NULL, horizon = 1,
  fixedwindow = TRUE, verboseIter = FALSE, returnData = TRUE,
  returnResamp = "final", savePredictions = FALSE,
  classProbs = FALSE, summaryFunction = defaultSummary,
  selectionFunction = "best", seeds = NA,
  preProcOptions = list(thresh = 0.95, ICAComp = 3, k = 5),
  sampling = NULL, index = NULL, indexOut = NULL,
  timingSamps = 0, predictionBounds = rep(FALSE, 2),
  adaptive = list(min = 5, alpha = 0.05, method = "gls",
  complete = TRUE), trim = FALSE, allowParallel = TRUE)

```

- 1) trainControl();
- 2) train.index;
- 3) preProcess();
- 4) conditional.

45. Метод \_\_\_\_\_, заданный объектом `trControl = trainControl()`, обеспечивает сканирование настраиваемых параметров и оценку эффективности модели на каждой итерации по определенным критериям качества. При построении каждой частной модели предварительно осуществляется предобработка данных с использованием методов, перечисленных в `preProcess` (и с учетом опций `preProcOptions` объекта `trControl`).

**(перекрестной проверки)**

46. \_\_\_\_\_ – общий метод, заключающийся в наложении дополнительных ограничений на искомые параметры, которые могут предотвратить излишнюю сложность модели. Смысл процедуры заключается в "стягивании" вектора коэффициентов  $\beta$  в ходе их настройки таким образом, чтобы они в среднем оказались несколько меньше по абсолютной величине, чем это было бы при оптимизации по МНК.

**(Регуляризация)**

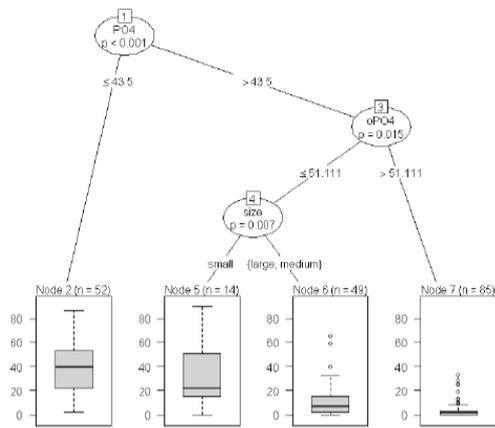
47. Реализация PLS-регрессии в среде R представлена в пакете \_\_\_\_\_, который включает в себя большое количество функций для построения регрессионных моделей, создания диагностических графиков и извлечения информации из результатов вычислений.

**(pls)**

48. Деревья решений осуществляют разбиение пространства в соответствии с некоторым набором правил разбиения.

**(верно/неверно)**

49. На графике показано:



- 1) дерево cpart без оптимизации параметра mincriterion;
- 2) дерево cpart после оптимизации параметра mincriterion;
- 3) дерево rpart с обрезанием ветвей;
- 4) дерево rpart без обрезания ветвей.

50. Стандартный механизм проверки статистических гипотез, который предотвращает переусложнение модели, реализован в функции

- 1) ctree;
- 2) scale;
- 3) plot;
- 4) street.

51. Метод \_\_\_\_\_ представляет собой дальнейшее улучшение бэггинга деревьев решений, которое заключается в устранении корреляции между деревьями. Смысл этой процедуры заключается в том, что с вероятностью  $(p-m)/p$  блокируется какой-нибудь потенциально доминирующий предиктор, стремящийся войти в каждое дерево. **(случайного леса)**

52. Методом улучшения предсказаний является \_\_\_\_\_, идея которого заключается в итеративном процессе последовательного построения частных моделей.

**(бустинг)**

53. Пакет \_\_\_\_\_ предоставляет инструментарий для описания, вычисления и графического изображения основных параметров последовательностей, в том числе оценки характеристик отдельных последовательностей (таких как длина, время пребывания, энтропия); расчета агрегированных характеристик совокупностей (скорости переходов, средней длительности пребывания в каждом состоянии, частоты повторения последовательностей) и пр.

**(TraMiner)**

54. Индекс сложности (complexity) последовательности рассчитывается как

- 1) геометрическое среднее энтропии и числа внутренних переходов, отнесенных к их максимальным значениям;
- 2) арифметическое среднее энтропии и числа внутренних переходов, отнесенных к их максимальным значениям;

3) квадратическое среднее энтропии и числа внутренних переходов, отнесенных к их максимальным значениям.

55. В основе \_\_\_\_\_ анализа лежит предположение том, что описания объектов каждого k-класса представляют собой реализации многомерной случайной величины, распределенной по нормальному закону

$$N_m(\mu_k; \Sigma_k) \quad \text{и ковариационной матрицей} \quad C_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (\mathbf{x}_{ik} - \mu_k)^T (\mathbf{x}_{ik} - \mu_k)$$

**(дискриминантного)**

56. Соотнесите определения с видами дискриминантного анализа

- |  |   |
|--|---|
| 1. Линейный дискриминантный анализ         | А) является своеобразным промежуточным звеном между LDA и QDA.  |
| 2. Квадратичный дискриминантный анализ     | Б) весьма эффективен, когда разделяющая поверхность между классами имеет ярко выраженный нелинейный характер; |
| 3. Регуляризованный дискриминантный анализ | В) Позволяет оценивать различия между двумя и более группами объектов по нескольким переменным одновременно;  |

Ответ: 1-В; 2-Б; 3-А.

57. Самый популярный в статистической литературе набор данных, часто используемый для иллюстрации работы различных алгоритмов классификации:

- 1) ирисы Фишера;
- 2) тюльпаны Фишера;
- 3) свечи Фишера;
- 4) нет верного ответа.

58. Основная функция \_\_\_\_\_ – сформировать выходной сигнал у в зависимости от сигналов x поступающих на его входы

**(искусственного нейрона)**

59. По способу соединения нейронов выделяют сети с разной архитектурой:

- 1) перцептроны;
- 2) сети адаптивного резонанса; рециркуляционные;
- 3) рекуррентные;
- 4) встречно распространения;
- 5) все ответы верные.

60. Для обучения ИНС в среде R используются два пакета – \_\_\_\_\_ .  
**(neuralnet и nnet)**

61. Центроиды это:

- 1) максимально удаленные друг от друга центры сгущений точек  $C_k$  с минимальным разбросом внутри каждого кластера;
- 2) минимально удаленные друг от друга центры сгущений точек  $C_k$  с минимальным разбросом внутри каждого кластера;
- 3) максимально удаленные друг от друга центры сгущений точек  $C_k$  с максимальным разбросом внутри каждого кластера;
- 4) минимально удаленные друг от друга центры сгущений точек  $C_k$  с максимальным разбросом внутри каждого кластера.

62. Функция \_\_\_\_\_ выполняет сканирование разбиений с различными значениями  $k$ , и строит график  $S_{\text{mean}}$  от  $k$ .

**(fviz\_nbclust())**

63. Эта функция может быть использована для визуализации результатов по методам  $k$ -средних, PAM, CLARA и Fanny. Ее простейший формат имеет вид:

```
fviz_cluster(object, data = NULL, stand = TRUE,
              geom = c("point", "text"),
              frame = TRUE, frame.type = "convex")
```

**(fviz\_cluster)**

64. Функции, которые используют в R для проведения иерархической классификации:

```
hclust(d, method = "complete")
agnes(x, metric = "euclidean", stand = FALSE, method = "average")
diana(x, metric = "euclidean", stand = FALSE)
```

где

- $d$ : матрица дистанций, полученная функцией `dist()` или как-то иначе;
- `method`: метод агломерации, определяемый одним из значений "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" или "centroid";
- $x$ : таблица данных для вычисления расстояний (наблюдения – по строкам, признаки – по столбцам);
- `metric`: метрика расстояний между наблюдениями, т.е. "euclidean" или "manhattan";
- `stand = TRUE`: осуществляется стандартизация каждой переменной таблицы (по столбцам).

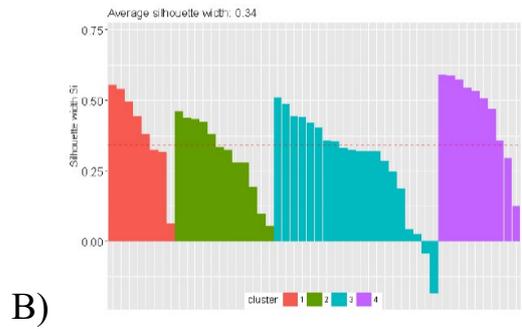
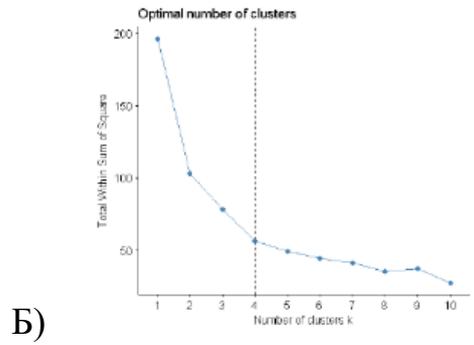
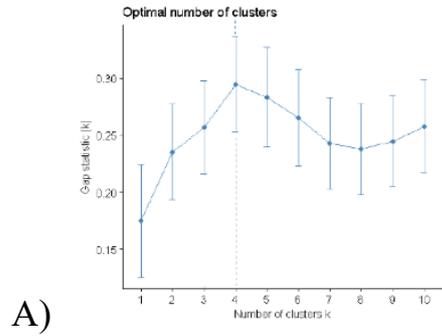
**(Верно/Неверно)**

65. В среде R кластеризация на главные компоненты реализована в пакете

- 1) FactoMiner;
- 2) trainControl();
- 2) train.index;
- 3) preprocess().

66. Соотнесите графики с методом выбора оптимального числа кластеров:

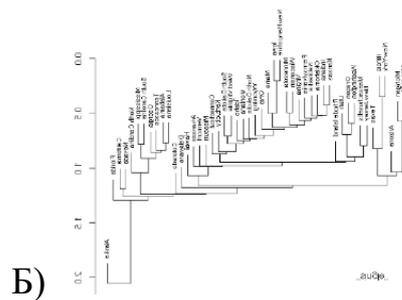
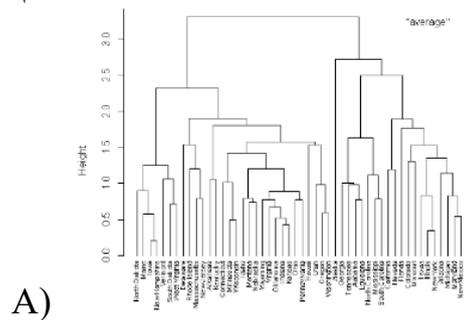
- 1) выбор оптимального числа кластеров по методу «локтя»:
- 2) выбор оптимального числа кластеров по методу статистики разрыва (GAP-статистики):
- 3) выбор оптимального числа кластеров по методу среднего силуэта:

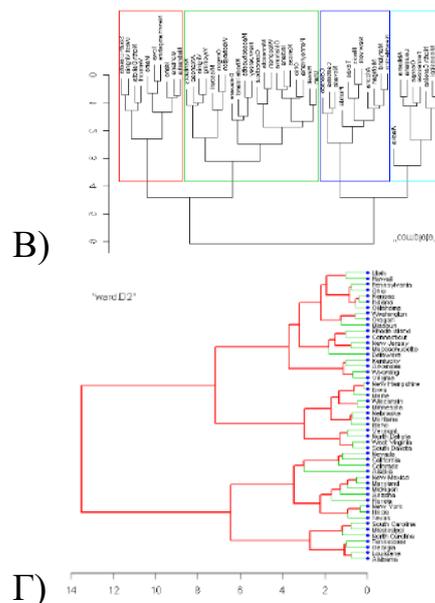


Ответ: 1-Б; 2-А; 3-В

67. Соотнесите графики с различными вариантами формирования дендрограмм и способами их визуализации:

- 1) метод средней связи:
- 2) метод одиночной связи:
- 3) метод полной связи:
- 4) метод Уорда:





Ответ: 1-А; 2-Б; 3-В; 4-Г.

68. Для количественной оценки различий между вариантами кластеризации могут быть использованы \_\_\_\_\_ или различные ранговые индексы.

- 1) кофенетическая корреляция;
- 2) каноническая корреляция;
- 3) кубическая корреляция;
- 4) геометрическая корреляция.

69. В среде R кластеризация на главные компоненты реализована в пакете \_\_\_\_\_.

**(Factominer)**

70. С использованием интерактивной оболочки Rattle во вкладке Model, можно построить такие модели классификации:

- 1) деревья решений;
- 2) случайные леса и бустинг;
- 3) машину опорных векторов;
- 4) линейную модель и нейронную сеть;
- 5) все ответы верные.

Критерии и шкала оценивания по оценочному средству «тесты»

Шкала оценивания (интервал баллов)	Критерий оценивания
5	Тесты выполнены на высоком уровне (правильные ответы даны на 90-100% тестов)
4	Тесты выполнены на среднем уровне (правильные ответы даны на 75-89% тестов)
3	Тесты выполнены на низком уровне (правильные ответы даны на

	50-74% тестов)
2	Тесты выполнены на неудовлетворительном уровне (правильные ответы даны менее чем на 50% тестов)

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

1. Основные понятия многомерных статистических методов.
2. Проблемы, решаемые с помощью многомерных статистических методов.
3. Практическое использование многомерных методов в статистическом анализе.
4. Реализация моделей Data Mining в среде R.
5. Основные шаги построения и верификации моделей.
6. Использование алгоритмов ресэмплинга для тестирования и оптимизации параметров моделей.
7. Модели для предсказания класса объектов.
8. Проецирование многомерных данных на плоскости.
9. Многомерный статистический анализ данных.
10. Методы кластеризации.
11. Универсальный интерфейс доступа к функциям машинного обучения в пакете caret.
12. Обнаружение и удаление "ненужных" предикторов.
13. Предварительная обработка данных: преобразование и групповая трансформация переменных.
14. Заполнение пропущенных значений в данных.
15. Функция train() пакета caret.
16. Селекция оптимального набора предикторов линейной модели.
17. Регуляризация, частные наименьшие квадраты и kNN-регрессия.
18. Построение деревьев регрессии.
19. Ансамбли моделей: бэггинг, случайный лес, бустинг.
20. Сравнение построенных моделей и оценка информативности предикторов
21. Деревья регрессии с многомерным откликом.
22. Классификация в бинарных пространствах с использованием классических моделей.
23. Бинарные деревья решений.
24. Поиск логических закономерностей в данных.
25. Алгоритмы выделения ассоциативных правил.
26. Анализ последовательностей знаков или событий.
27. Дискриминантный анализ.
28. Метод опорных векторов.
29. Ядерные функции машины опорных векторов.
30. Деревья классификации, случайный лес и логистическая регрессия. Процедуры сравнения эффективности моделей классификации.

31. Ирисы Фишера и метод k-ближайших соседей.
32. Наивный классификатор Байеса.
33. Классификация в линейном дискриминантном пространстве.
34. Нелинейные классификаторы в R.
35. Модель мультиномиального логита.
36. Классификаторы на основе искусственных нейронных сетей.
37. Модель логита для порядковой переменной.
38. Настройка параметров нейронных сетей средствами пакета caret.
39. Методы комплексации модельных прогнозов.
40. Обобщенные линейные модели для счетных данных.
41. ZIP- и барьерные модели счетных данных
42. Преобразование данных и вычисление матрицы расстояний.
43. Непараметрический дисперсионный анализ матриц дистанций.
44. Методы ординации объектов и переменных: построение и сравнение диаграмм.
45. Оценка связи ординации с внешними факторами.
46. Неметрическое многомерное шкалирование и построение распределения чувствительности видов.
47. Алгоритмы кластеризации, основанные на разделении.
48. Иерархическая кластеризация.
49. Оценка качества кластеризации.
50. Другие алгоритмы кластеризации.
51. Самоорганизующиеся карты Кохонена.
52. Начало работы с пакетом rattle.
53. Описательная статистика и визуализация данных.
54. Построение и тестирование моделей классификации.
55. Deskриптивные модели (обучение без учителя).

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

Шкала оценивания	Характеристика знания предмета и ответов
отлично (5)	Студент глубоко и в полном объеме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач.
хорошо (4)	Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.
удовлетворительно (3)	Студент знает только основной программный материал, допускает неточности, недостаточно четкие формулировки, непоследовательность в ответах, излагаемых в устной или

	<p>письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.</p>
<p>неудовлетворительно (2)</p>	<p>Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.</p>

### Лист изменений и дополнений

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