

**МИНОБРНАУКИ РОССИИ**

**Федеральное государственное бюджетное образовательное учреждение  
высшего образования**

**«Чувашский государственный университет имени И. Н. Ульянова»  
(ФГБОУ ВО «ЧГУ им. И.Н. Ульянова»)**

Факультет информатики и вычислительной техники

Кафедра компьютерных технологий

Утверждено  
на заседании кафедры компьютерных  
технологий  
Заведующий кафедрой Т. А. Лавина

 \_\_\_\_\_ 25.03.2022

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
(ФОНД ОЦЕНОЧНЫХ СРЕДСТВ)**

***«Методы анализа данных в бизнес-аналитике»***

Направление подготовки / специальность 09.04.03 Прикладная информатика

Квалификация выпускника Магистр

Направленность (профиль) / специализация « Искусственный интеллект и бизнес-аналитика»

Год начала подготовки - 2022

Составитель(и):

Доцент, кандидат технических наук Ржавин В.В.

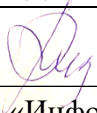
Согласовано  
методической комиссией факультета информатики и вычислительной техники  
25.03.2022, протокол № 8

Декан факультета А. В. Щипцова

**Паспорт**  
оценочных материалов для проведения текущего контроля и  
промежуточной аттестации обучающихся по дисциплине (модулю)  
Методы анализа данных в бизнес-аналитике

Перечень оценочных материалов и индикаторов достижения компетенций,  
сформированность которых они контролируют

Наименование оценочного средства	Коды индикаторов достижения формируемых компетенции	Номер приложения
Тест	ИД-2 <small>ОПК-1</small> ИД-3 <small>ОПК-1</small> ИД-1 <small>ОПК-4</small> ИД-2 <small>ОПК-4</small> ИД-3 <small>ОПК-4</small>	1
Выполнение лабораторных работ	ИД-1 <small>ОПК-1</small> ИД-2 <small>ОПК-1</small> ИД-3 <small>ОПК-1</small> ИД-1 <small>ОПК-4</small> ИД-2 <small>ОПК-4</small> ИД-3 <small>ОПК-4</small>	2
Собеседование	ИД-1 <small>ОПК-1</small>	3
Экзамен	ИД-1 <small>ОПК-1</small> ИД-2 <small>ОПК-1</small> ИД-3 <small>ОПК-1</small> ИД-1 <small>ОПК-4</small> ИД-2 <small>ОПК-4</small> ИД-3 <small>ОПК-4</small>	4

Разработал: \_\_\_\_\_  В.В. Ржавин  
Утверждено на заседании кафедры «Информационные системы»  
протокол № 3 от «11» октября 2021 года  
Заведующий кафедрой \_\_\_\_\_ А.А. Романов

## I. Текущий контроль

## Приложение 1

### Тесты

#### 1. Процедура проведения тестирования

Количество проводимых тестов в течение всего периода освоения дисциплины	4 теста
Общее количество тестовых вопросов в банке тестов	10 вопросов
Количество задаваемых тестовых вопросов в одном тесте	5 вопросов
Формат проведения тестирования	Электронный
Сроки / Периодичность проведения тестирования	На лабораторных занятиях 4 раза в семестр

#### 2. Шкала оценивания с учетом срока сдачи

Процент правильных ответов	Балл
90 % правильных ответов;	Отлично
75%-89% правильных ответов;	Хорошо
51%-74% правильных ответов;	Удовлетворительно
менее 51% правильных ответов.	Неудовлетворительно

#### 3. Тестовые задания

##### Тестовые задания по теме «Введение в бизнес-аналитику. Предварительный анализ данных»

1. Как называется совокупность подходов и методов для автоматизации анализа текстовых документов, включая задачи установления степени сходства текстов, категоризации документов, формирования аннотаций и пр.?
  1. Генетические алгоритмы (Genetic algorithms)
  2. Вычислительный интеллект (Computational intelligence)
  3. Анализ текстов (Text analysis)
  4. Кластер-анализ (Cluster analysis)
2. К классу описательных задач относятся:
  1. кластеризация и классификация;
  2. кластеризация и поиск ассоциативных правил;
  3. классификация и регрессия;
  4. классификация и поиск ассоциативных правил.
3. К классу предсказательных задач относятся:
  1. кластеризация и классификация;
  2. кластеризация и поиск ассоциативных правил;
  3. классификация и регрессия;
  4. классификация и поиск ассоциативных правил.
4. К классу задач supervised learning (обучение с учителем) относятся:
  1. кластеризация и классификация;
  2. кластеризация и поиск ассоциативных правил;
  3. классификация и регрессия;
  4. классификация и поиск ассоциативных правил.

5. К классу задач *unsupervised learning* (обучение без учителя) относятся:
  1. кластеризация и классификация;
  2. кластеризация и поиск ассоциативных правил;
  3. классификация и регрессия;
  4. классификация и поиск ассоциативных правил.
6. Задача классификации сводится к ...
  1. нахождения частых зависимостей между объектами или событиями;
  2. определения класса объекта по его характеристикам;
  3. определение по известным характеристикам объекта значение некоторого его параметра;
  4. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
7. Задача регрессии сводится к ...
  1. нахождения частых зависимостей между объектами или событиями;
  2. определения класса объекта по его характеристиками;
  3. определение по известным характеристикам объекта значение некоторого его параметра;
  4. поиска независимых групп и их характеристик в всем множестве анализируемых данных.
8. Задача кластеризации заключается в ...
  1. нахождения частых зависимостей между объектами или событиями;
  2. определения класса объекта по его характеристиками;
  3. определение по известным характеристиками объекта значение некоторого его параметра;
  4. поиска независимых групп и их характеристик в всем множестве анализируемых данных.
9. Целью поиска ассоциативных правил является ...
  1. нахождения частых зависимостей между объектами или событиями;
  2. определения класса объекта по его характеристиками;
  3. определение по известным характеристиками объекта значение некоторого его параметра;
  4. поиска независимых групп и их характеристик в всем множестве анализируемых данных.
10. Модели классификации описывают ...
  1. правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
  2. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
  3. функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
  4. группы, на которые можно разделить объекты, данные о которых подвергаются анализа.

#### **Тестовые задания по теме «Исследование зависимостей»**

1. Регрессивные модели описывают ...
  1. правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
  2. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
  3. функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
  4. группы, на которые можно разделить объекты, данные о которых подвергаются анализа

2. Задача регрессии сводится к ...
  1. нахождения частых зависимостей между объектами или событиями;
  2. определения класса объекта по его характеристикам;
  3. определение по известным характеристикам объекта значение некоторого его параметра;
  4. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
3. Многомерный анализ – это:
  1. одновременный анализ по нескольким измерениям;
  2. одновременный анализ по нескольким параметрам;
  3. одновременный анализ по нескольким данным.
4. Коэффициент парной корреляции между процентом охвата населения прививками и заболеваемостью на 10 000 населения равен (-0,86). Можно сделать следующие выводы:
  1. связь между изучаемыми явлениями отсутствует, т.к. коэффициент корреляции отрицательный
  2. связь между изучаемыми явлениями обратная (отрицательная)
  3. связь между изучаемыми явлениями сильная и обратная
  4. связь между изучаемыми явлениями слабая и обратная
  5. связь между изучаемыми явлениями средняя и обратная.
5. При каком значении коэффициента корреляции связь можно считать умеренной?
  1.  $r = 0,47$ ;
  2.  $r = 0,71$ .
  3.  $r = 0,21$ ;
6. Верно или нет следующее утверждение (и почему): если линия регрессии имеет большой наклон, то корреляция между переменными также должна быть большой?
  1. Да, потому что при большем наклоне увеличение одной переменной ведет к более сильному увеличению другой переменной
  2. Да, потому что коэффициент корреляции пропорционален углу наклона линии регрессии.
  3. Нет, корреляция не зависит угла наклона.
7. Определить коэффициент корреляции двух переменных: агрессивности ( $x_a$ ) и IQ у школьников ( $y_{iq}$ ) по полученным данным тестирования

№	$X_{agr}$	$Y_{iq}$
1	24	100
2	27	115
3	26	117
4	21	119
5	20	134
6	31	94
7	26	105
8	22	103
9	20	111
10	18	124
11	30	122
12	29	109
13	24	110
14	26	86

1. 0,27
  2. -0,42
  3. -0,64
  4. 0,81
8. Знания 10 студентов проверены по двум тестам А и В. Оценки по стобалльной системе сведены в таблицу

А	95	90	86	84	75	70	62	60	57	50
В	92	93	83	80	55	60	45	72	62	70

Найти выборочный коэффициент ранговой корреляции Спирмена между оценками по двум тестам.

1. 0,27
2. 0,42
3. 0,64
4. 0,81

9. По данным задачи №8 определить коэффициент ранговой корреляции Кендалла
1. 0,28
  2. 0,47
  3. 0,63
  4. 0,81
10. Если величина коэффициента корреляции Пирсона между переменными равна  $-0,75$  то
1. это очень слабая корреляция и в большинстве случаев мы не берем ее в расчет;
  2. это высокая корреляция и мы принимаем ее в расчет.
11. Что означает совсем низкое или нулевое значение коэффициента корреляции двух количественных признаков?
1. наличие неизвестного вида связи
  2. наличие квадратичной зависимости
  3. отсутствие линейной связи
  4. наличие линейной связи
12. Термин регрессия в статистике понимают как: а) функцию связи, зависимости; б) направление развития явления вспять; в) функцию анализа случайных событий во времени; г) уравнение линии связи
1. а, б
  2. в, г
  3. а, г
13. Для анализа связи между двумя номинальными признаками составляют следующую таблицу. Строки таблицы соответствуют категориям одного признака, а столбцы — категориям другого признака. Элемент на пересечении строки и столбца — количество объектов, обладающих соответствующими категориями и того и другого признаков. Такая таблица называется таблицей \_\_\_\_\_
14. Рассмотрим модель «мешок слов» в задаче выявления корреляции по данным следующей таблицы.

Статья	Ключевые слова									
	пить	равный	греть	играть	легкий	цена	свобода	талант	налог	женский
F1	1	2	0	1	2	0	0	0	0	2
F2	0	0	0	1	0	1	0	2	0	2
F3	0	2	0	0	0	0	0	1	0	2
F4	2	1	0	0	0	2	0	2	0	1
E1	2	0	1	2	2	0	0	1	0	0
E2	0	1	0	3	2	1	2	0	0	0
E3	1	0	2	0	1	1	0	3	1	1
E4	0	1	0	1	1	0	1	1	0	0
N1	0	0	2	0	1	2	0	0	2	0
N2	1	0	2	2	0	2	2	0	0	0
N3	0	0	1	1	2	1	1	0	2	0
N4	0	0	1	0	0	2	2	0	2	0

Объектами являются газетные статьи, разделенные на три категории в соответствии с темами «Феминизм», «Развлечения» и «Домохозяйство». Каждая статья характеризуется своим набором ключевых слов, представленных в соответствующей строке таблицы. Чтобы уменьшить эффект случайности отбора статей в таблицу данных, примем, что «мешок» содержит по одному появлению каждого ключевого слова, независимо от того, появилось ли оно в статьях данной категории или нет. Какова вероятность слова «играть» в категории N?

1. 2/41
2. 7/41
3. 4/41
4. 3/41

#### Тестовые задания по теме «Снижение размерности признаков пространства»

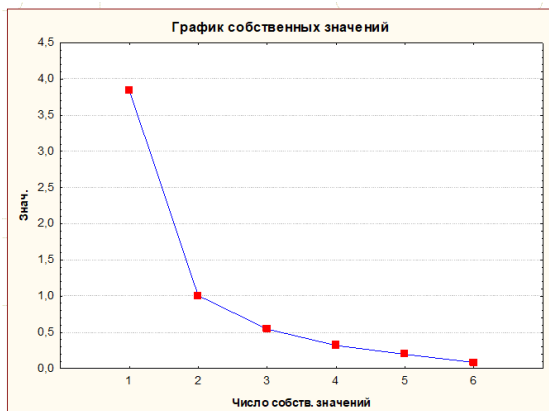
1. Укажите действия, позволяющих уменьшить число факторов:
  1. Устранение дублирующей информации при наличии сильно коррелированных признаков
  2. Редукция слабоинформативных (маломеняющихся для различных объектов) признаков
  3. Агрегирование (объединение) нескольких признаков в один





2. Математический метод нахождения главных осей заключается в
  1. вычислении собственных чисел и собственных векторов ковариационной матрицы  $S$ .
  2. вычислении детерминанта ковариационной матрицы  $S$
3. Даны четыре примера (наблюдения) в трехмерном пространстве признаков:  $A(1;4;10)$ ,  $B(2;5;6)$ ,  $C(1;3;8)$  и  $D(2;4;8)$ . В результате применения метода главных компонент исходное пространство признаков свели к двумерному пространству признаков на плоскости. Найдите евклидово расстояние между примерами  $C$  и  $D$  в редуцированном пространстве с точностью до одного знака после запятой: \_\_\_\_\_
4. Какова идея метода главных компонент?
  - 1) поиск гиперплоскости заданной размерности, такой чтобы ошибка проектирования выборки на данную гиперплоскость была минимальной
  - 2) поиск проекции на гиперплоскость с сохранением большей части дисперсии в данных
  - 3) проекция данных на гиперплоскость с критической ошибкой проектирования
5. Укажите верное утверждение
  1. Метод главных компонент использует меньшее количество компонент, в отличие от метода независимых компонент
  2. Метод главных компонент добивается ортогональности между полученными компонентами, а метод независимых компонент - не ортогональности
  3. Метод независимых компонент работает с коррелированными данными, в отличие от метода главных компонент
  4. Метод главных компонент применяется в основном для задач, где необходимо разделять сигналы, а метод независимых компонент - для визуального разделения данных
6. Каковы недостатки метода главных компонент?
  1. координаты объектов в новом пространстве определены не однозначно
  2. проблема с вычислением собственных векторов ковариационной матрицы, при большом количестве данных
  3. существует произвол в выборе координат объектов в новом пространстве
  4. общая сложность алгоритма
7. Выберите сферы применения метода главных компонент
  1. Визуализация данных
  2. Построение деревьев решений
  3. Обработка изображений
  4. Выявление максимальной избыточности
  5. Отбор признаков
8. Укажите достоинство использования метода главных компонент
  1. Простой алгоритм
  2. Координаты объектов в новом пространстве определены однозначно
  3. Легкость с вычислением собственных векторов ковариационной матрицы в случае большого количества данных
  4. все перечисленное

9. Согласно критерию каменной осыпи число общих факторов равно \_\_\_\_\_



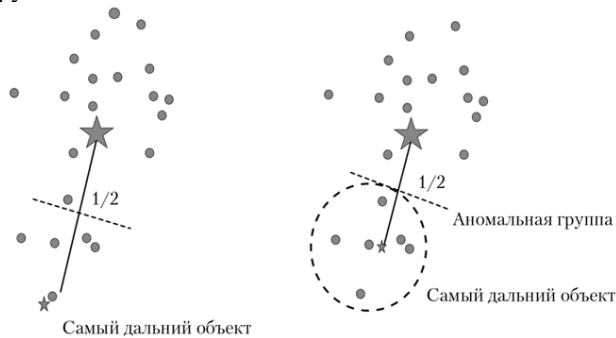
10. По критерию Кайзера отбираются только факторы с собственными значениями

1. равными или большими 1
2. меньшими 1
3. меньше натурального логарифма

#### Тестовые задания по теме «Классификация многомерных наблюдений»

1. Кластеризация — ...
  1. это установление зависимости непрерывной выходной переменной от входных переменных
  2. эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
  3. выявление закономерностей между связанными событиями
  4. это установление зависимости дискретной выходной переменной от входных переменных.
2. Задача кластеризации заключается в ...
  1. нахождения частых зависимостей между объектами или событиями;
  2. определения класса объекта по его характеристикам;
  3. определение по известным характеристикам объекта значение некоторого его параметра;
  4. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
3. Расставьте в правильном порядке шаги алгоритма К-средних
  1. Предобработка
  2. Инициализация аномального центра
  3. Обновление аномального центра
  4. Обновление аномальной группы
  5. Выдача результатов
4. Поставьте в правильном порядке шаги алгоритма uK-средних ( $t$ ), где  $t$  — порог разрешения, т. е. задаваемое пользователем минимальное количество объектов в аномальной группе, необходимое, чтобы она могла восприниматься как генератор отдельного кластера.
  1. Условие останова
  2. Метод К-средних
  3. Настройка
  4. Аномальная группа
  5. Отбрасывание малых кластеров

5. На рисунке изображены первая (слева) и финальная (справа) итерации извлечения аномальной группы из структуры некоторого множества объектов. Малая звезда представляет центр аномальной группы.



Как называется точка, обозначенная большой звездой?

1. центральная точка
  2. реперная точка
  3. нормальная точка
  4. аномальная точка
6. Спорный объект кластеризации — это объект, который по мере сходства ...
1. может быть отнесен к нескольким кластерам
  2. не может быть отнесен ни к одному кластеру
  3. может быть отнесен более чем к двум кластерам
7. Процедура, которая приводит значения всех преобразованных переменных к единому диапазону значений путем выражения через отношение этих значений к некоей величине, отражающей определенные свойства, это — ...
1. стандартизация
  2. нормирование
  3. оба ответа верны
8. Работа кластерного анализа опирается на следующие предположения (выберите неверный ответ):
1. рассматриваемые признаки объекта в принципе допускают желательное разбиение объектов на кластеры
  2. правильность выбора масштаба или единиц измерения признаков
  3. отнесение всех объектов к одному из predetermined признаков
9. Иерархические агломеративные методы характеризуются ...
1. последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
  2. делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп
  3. сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
10. Объект относится к кластеру, если ...
1. расстояние от объекта до центра кластера меньше радиуса кластера
  2. расстояние от объекта до центра кластера меньше диаметра кластера
  3. расстояние от объекта до центра кластера больше радиуса кластера
11. Иерархические дивизимные методы характеризуются ...
1. последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
  2. делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп
  3. сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
12. При применении кластерного анализа переменные ...
1. должны измеряться в сравнимых шкалах
  2. могут измеряться в каких угодно шкалах
  3. должны быть только числовыми

13. Характеристикой каких групп методов являются последовательное объединение исходных элементов и соответствующее уменьшение числа кластеров?
1. иерархические агломеративные методы
  2. иерархические дивизимные (делимые) методы
  3. и тех, и других
14. Деление одного кластера на меньшие кластеры, в результате чего образуется последовательность расщепляющих групп. Характеристика каких групп методов описана выше?
1. иерархические агломеративные методы
  2. иерархические дивизимные (делимые) методы
  3. и тех, и других

## Выполнение лабораторных работ

## 1. Процедура выполнения лабораторных работ

Количество проводимых лабораторных работ в течение всего периода освоения дисциплины	4 работы
Формат проведения результатов	Бумажный

## 2. Шкала оценивания с учетом срока сдачи

Процент правильных ответов	Балл
90 % правильных ответов;	отлично
75%-89% правильных ответов;	хорошо
51%-74% правильных ответов;	удовлетворительно
менее 51% правильных ответов.	неудовлетворительно

Оценка «зачтено» проставляется студенту, выполнившему и защитившему в полном объеме лабораторные работы в течение семестра, чей уровень знаний, умений и навыков соответствует уровню оценок «отлично», «хорошо» или «удовлетворительно».

Оценка «не зачтено» проставляется студенту, не выполнившему и (или) не защитившему в полном объеме лабораторные работы в течение семестра, либо чей уровень знаний, умений и навыков соответствует уровню оценки «неудовлетворительно».

## 3. Перечень лабораторных работ

1. Ввод и предварительный анализ данных
2. Регрессионный анализ
3. Факторный анализ
4. Кластерный анализ

Лабораторные работы (электронная версия) представлены в модульной объектно-ориентированной динамической учебной среде (MOODLE) университета.

## Собеседование

## 1. Процедура проведения

Тип собеседования	По лабораторным работам
Общее количество вопросов для собеседования	3-6 вопросов
Количество основных задаваемых при собеседовании вопросов	2 вопроса
Формат проведения собеседования	Устно
Периодичность проведения собеседования	На лабораторных занятиях в течение семестра (раз в 4 недели)
Методические рекомендации (при необходимости)	

## 2. Шкала оценивания с учетом срока сдачи

Критерии оценивания	Балл
наличие глубоких и исчерпывающих знаний в объёме пройденного программного материала правильные и уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, знание дополнительно рекомендованной литературы	отлично
наличие твердых и достаточно полных знаний программного материала, незначительные ошибки при освещении заданных вопросов, правильны действия по применению знаний на практике, четкое изложение материала	хорошо
наличие твердых знаний пройденного материала, изложение ответов с ошибками, уверенно исправляемыми после дополнительных вопросов, необходимость наводящих вопросов, правильные действия по применению знаний на практике	удовлетворительно
наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы	неудовлетворительно

## 3. Примерный перечень вопросов для собеседования

1. Перечислите основные задачи, решаемые в анализе данных
2. Назовите этапы решения задач. Дайте им краткую характеристику.
3. Приведите классификацию статистических переменных
4. Как оценивается связь порядковых переменных?
5. Назовите этапы регрессионного анализа.
6. Как строится таблица сопряженности?
7. Как производится выбор модели регрессии?
8. Как проводится оценка параметров модели?
9. Как проверяется адекватность модели регрессии?

10. В чем заключаются необходимость снижения размерности признакового пространства?
11. Каковы предпосылки, обуславливающие возможность снижения размерности признакового пространства?
12. Что понимается под «мерой информативности» в методе главных компонент?
13. Сформулировать определение  $k$ -ой главной компоненты
14. В чем заключается алгоритм нахождения коэффициентов линейного преобразования исходных признаков?
15. Каковы основные числовые характеристики главных компонент?
16. Каковы свойства матрицы факторных нагрузок?
17. Чему равен вклад главных компонент в дисперсию  $i$ -го признака?
18. Чему равен вклад  $j$ -ой главной компоненты в суммарную дисперсию признаков?
19. Как определить матрицу индивидуальных значений главных компонент?
20. Описать процедуру статистического анализа при реализации метода главных компонент
21. Привести математическую модель главных факторов
22. Что представляют собой составляющие в разложении дисперсии  $i$ -го признака?
23. В чем заключается фундаментальная теорема факторного анализа?
24. Определить понятие факторной структуры
25. Определить понятия пространства главных факторов и полного факторного пространства
26. В чем заключается алгоритм метода главных факторов?
27. Дать понятия простой ортогональной, косоугольной и случайной структуры
28. В чем идея и суть вращения пространства главных факторов?
29. Какие существуют методы оценки индивидуальных значений главных факторов?
30. Описать процедуру статистического анализа при реализации метода главных факторов
31. Чему равны вклад и относительный вклад третьей главной компоненты в суммарную дисперсию исходных признаков?
32. Каков уровень информативности первых трех главных компонент?
33. Сколько главных компонент нужно построить для достижения уровня информативности 100%?
34. Каким образом на основании собственных векторов рассчитаны коэффициенты линейного преобразования исходных признаков?
35. Чему равен коэффициент корреляции между первой главной компонентой и четвертым исходным признаком?
36. Ранжируйте исходные признаки по убыванию удельного веса их влияния на формирование первой главной компоненты
37. Каким образом дается интерпретация главных компонент?
38. Назовите этапы кластерного анализа.
39. Как вычисляется расстояние между объектами кластерного анализа?
40. Как вычисляется расстояние между классами кластерного анализа?

Экзамен (зачет с оценкой)

1. Процедура проведения

Общее количество вопросов к экзамену (зачету с оценкой)	44 вопроса
Количество вопросов в билете	2 вопроса
Наличие задач в билете	Нет
Формат проведения	Устно

2. Шкала оценивания с учетом текущего контроля работы обучающегося в семестре

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
обучающийся показал глубокие знания материала по поставленному вопросу, грамотно, логично и стройно его излагает	Отлично
обучающийся твердо знает материал, грамотно его излагает, но допускает несущественные неточности в ответе на вопрос	Хорошо
обучающийся показывает знания только основных положений по поставленному вопросу, требует в отдельных случаях наводящих вопросов для принятия правильного решения, допускает отдельные неточности	Удовлетворительно
обучающийся допускает грубые ошибки в ответе на поставленный вопрос	Неудовлетворительно

3. Вопросы к экзамену

1. Основные задачи, решаемые в анализе данных.
2. Этапы решения задач. Постановка задачи. Ввод данных в обработку.
3. Этапы решения задач. Качественный анализ. Количественное описание данных.

Интерпретация результатов.

4. Математические модели в прикладной статистике.
5. Статистические модели ряда наблюдений. Модель выборки.
6. Статистические модели ряда наблюдений. Модель тренда с ошибкой. Модель случайного процесса.
7. Описательные характеристики. Математические ожидания и моменты случайных величин. Характеристики центра группирования значений случайной величины
8. Характеристики степени рассеяния случайной величины.
9. Многомерные случайные величины. Моменты второго порядка случайной величины.
10. Многомерные случайные величины. Выборочные ковариация и корреляция.



11. Введение в многомерный стохастический анализ. Исследование зависимостей. Основные прикладные цели.
12. Классификация статистических переменных. Виды статистического анализа.
13. Этапы корреляционного анализа. Причинный и статистический характер связи. Виды характеристик связи. Природа статистической взаимосвязи.
14. Количественные характеристики статистической взаимосвязи. Диаграммы рассеивания. Ковариация.
15. Коэффициент парной корреляции. Корреляционная матрица.
16. Частные коэффициенты корреляции. Коэффициент множественной корреляции
17. Корреляционная функция.
18. Характеристики многомерной статистической связи.
19. Связь порядковых переменных. Коэффициент Спирмена. Коэффициент Кендалла.
20. Анализ связей между классификационными переменными. Таблицы сопряженности. Проверка гипотезы о наличии связи.
21. Регрессионный анализ. Общая постановка задачи. Основные этапы регрессионного анализа
22. Выбор модели регрессии.
23. Уравнение регрессии. Оценка параметров модели
24. Анализ регрессии. Адекватность модели. Дисперсия адекватности и дисперсия воспроизводимости.
25. Проверка значимости коэффициентов регрессионной модели.
26. Исследование регрессионных остатков
27. Оценка точности регрессионной модели.
28. Постановка задачи МГК
29. Вычисление главных компонент.
30. Факторный анализ (ФА). Общая постановка задачи. Цели ФА.
31. Порядок выполнения факторного анализа. Основные этапы факторного анализа.
32. Определение главных факторов – факторизация. Задание числа факторов.
33. Факторный анализ. Вращение факторов.
34. Факторный анализ. Оценка значений общих факторов.
35. Кластерный анализ (КА) Общая постановка задачи. Этапы кластерного анализа.
36. Кластерный анализ. Меры сходства. Расстояние между отдельными объектами. Расстояние между классами в кластерном анализе.
37. Обзор методов кластерного анализа.
38. Методы кластерного анализа. Иерархические методы.
39. Методы кластерного анализа. Иерархические методы.
40. Методы кластерного анализа. Итеративные методы. Алгоритм k-средних.
41. Методы кластерного анализа. Итеративные методы. Алгоритм k-средних.
42. Методы кластерного анализа. Иерархические методы.
43. Методы кластерного анализа. Итеративные методы. Алгоритм k-средних.
44. Методы ассоциативного поиска.