

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

Федеральное государственное образовательное учреждение высшего образования

«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет управления

Кафедра математических и естественнонаучных дисциплин

Учебно-методический комплекс по дисциплине

Анализ данных

Образовательная программа

38.03.05 Бизнес-информатика

Профиль подготовки

Технологическое предпринимательство

Уровень высшего образования

бакалавриат

Форма обучения __очная

Статус дисциплины: _вариативная

Махачкала 2017

Рабочая программа дисциплины «Анализ данных» составлена в 2017 году в соответствии с требованиями ФГОС ВО по направлению подготовки 38.03.05 – Бизнес-информатика (бакалавриат) от «14» января 2010г. №27.

Разработчик: кафедра МиЕНД, проф. Муртузалиев М.М.
(кафедра, ФИО, ученая степень, ученое звание)

Рабочая программа дисциплины одобрена:

на заседании МиЕНД от «03» мая 2017 г., протокол № 8

Зав. кафедрой  Омарова Н.О.

(подпись)

на заседании Методической комиссии факультета управления от «12» июн 2017 г., протокол № 9.

Председатель  Камалова Т.А.

(подпись)

Рабочая программа дисциплины согласована с учебно-методическим управлением «12» 05 2017 г. 

(подпись)

Аннотация рабочей программы дисциплины

Дисциплина «Анализ данных» входит в вариативную часть по выбору образовательной программы бакалавриата по направлению 38.03.05 Бизнес-информатика.

Дисциплина реализуется на факультете управления кафедрой «Математических и естественнонаучных дисциплин».

Содержание дисциплины охватывает круг вопросов, связанных с формированием у студентов фундаментальных знаний и практических навыков, позволяющих использовать современные математико-статистические методы и программные средства для обработки и анализа данных. Дисциплина нацелена на формирование следующих компетенций выпускника: профессиональных – ОК-16, ПК-19; ПК-20.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: лекции, практические занятия, лабораторных занятий, самостоятельная работа.

Рабочая программа дисциплины предусматривает проведение следующих видов контроля: текущий контроль успеваемости в форме опросов, рефератов, дискуссий, тестов, решения задач и промежуточный контроль в форме зачета.

На изучение дисциплины отводится 18 часов лекций, 16 часов практических, 16 часов лабораторных занятий в 4 семестре, форма контроля – зачет.

Семестр	Учебные занятия						Форма промежуточной аттестации (зачет, дифференцированный зачет, экзамен)	
	в том числе							
	Всего	из них					СРС, в том числе экзамен	
		Лекции	Лабораторные занятия	Практические занятия	КСР	консультации		
4	108	18	16	16	-	-	58	Зачет

1. Цели освоения дисциплины

Целью преподавания дисциплины «Анализ данных» является приобретения студентами знания современных концепций построения и перспектив использования современных математико-статистических методов и программных средств для обработки и анализа данных.

Изучение данной дисциплины подготавливает студентов к освоению современных математико-статистических методов и программных средств для обработки и анализа данных, связанных с их будущей деятельностью.

Задачи изучения дисциплины включают:

- ознакомить студентов с методикой и современными методами анализа данных;
- научить обобщать и анализировать информацию;
- дать навыки применения различных методов анализа данных.

2. Место дисциплины в структуре ООП бакалавриата

Дисциплина «Анализ данных» входит в вариативную часть образовательной программы бакалавриата 38.03.05 (080500.62) – Бизнес-информатика.

Преподавание дисциплины «Анализ данных» основано на дисциплинах - линейная алгебра, теория вероятности и математическая статистика, эконометрика, макро- и микроэкономика, исследование операций.

В свою очередь она создаёт необходимые предпосылки для освоения программ таких дисциплин, как общая теория систем, рынки ИКТ и организация продаж, маркетинг, стратегический и инновационный менеджмент, дипломное проектирование и ряда дисциплин по выбору студента.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения).

Компетенции	Формулировка компетенции из ФГОС ВО	Планируемые результаты обучения (показатели достижения заданного уровня освоения компетенций)
ОК-16	способен работать с информацией	Знать: основные

	из различных источников	<p>принципы сбора и подготовки исходных данных;</p> <p>Уметь: систематизировать и обобщать информацию</p> <p>Владеть: основными приемами сбора информации и инструментальными средствами для обработки информации</p>
ПК-19	использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования	<p>Знать: основные методы анализа данных</p> <p>Уметь: применять современные математико-статистические методы и инструментальные средства для исследования объектов профессиональной деятельности</p> <p>Владеть: методами статистического анализа и прогнозирования</p>
ПК-20	использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования	<p>Знать: основные методы анализа данных</p> <p>Уметь: использовать математические, инструментальные средства для обработки, анализа и систематизации информации по теме исследования</p> <p>Владеть: навыками проведения факторного и кластерного анализа</p>

1.4. Структура и содержание дисциплины:

Общая трудоемкость дисциплины составляет 3 зачетные единицы, 108 часов.

Наименование тем	Количество часов					Формы текущего контроля успеваемости (по неделям семестра)	Форма промежуточной аттестации (по семестрам)
	в том числе по видам учебных занятий						
	Всего	Лекции	Практические занятия	Лабораторные	Самостоятельная работа		
1	2	3	4	5			
Модуль 1. Технологии обеспечения качественного анализа данных							
Тема 1. Технологии анализа данных.	6	2	0	0	4	Оценка по сто балльной шкале	
Тема 2. Консолидация данных.	10	2	2	2	4	Оценка по сто балльной шкале	
Тема 3. Трансформация данных.	10	2	2	2	4	Оценка по сто балльной шкале	
Тема 4. Визуализация данных.	10	2	2	2	4	Оценка по сто балльной шкале	
Итого по 1 модулю:	36	8	6	6	16	Контрольная работа	

Модуль 2. Инструменты Data mining

Тема 5. Оценка качества, очистка и предобработ ка данных.	12	2	2	2	6	Оценка по сто балльной шкале
Тема 6. Поиск ассоциативн ых правил.	12	2	2	2	6	Оценка по сто балльной шкале
Тема 7. Кластеризац ия. Тема 8. Классифика ция и регрессия. Статистичес кие методы.	12	2	2	2	6	Оценка по сто балльной шкале
Итого по 2- модулю:	36	6	6	6	18	Контрольна я работа

Модуль 3. Анализ и прогнозирование временных рядов.

Тема 9. Классифика ция и регрессия. Машинное обучение. Тема 10. Анализ и прогнозиров ание временных рядов.	18	2	2	2	12	Оценка по сто балльной шкале
Тема 11. Ансамбли моделей. Тема 12. Сравнение моделей.	18	2	2	2	12	Оценка по сто балльной шкале
Итого по 3-	36	4	4	4	24	Контрольна

модулю:						я работа
ИТОГО:	108	18	16	16	58	Зачет

2.2. Содержание курса

2.2. Лекционный материал по дисциплине.

Модуль 1. Технологии обеспечения качественного анализа данных

Тема 1. Технологии анализа данных.

Аналитический и информационный походы к моделированию. Формы представления, типы и виды анализируемых данных. Источники данных для анализа. Технология KDD (Knowledge Discovery in Databases). Введение в технологию Data Mining. Программный инструментарий анализа данных.

Тема 2. Консолидация данных.

Введение в консолидацию данных. Общая характеристика OLTP-систем. Предпосылки появления систем поддержки принятия решений (DSS – Decision Support System). Введение в хранилища данных (DW – Data Warehouse). Реляционные хранилища данных (ROLAP – Relational OLAP). Многомерные хранилища данных (MOLAP – Multidimensional OLAP). Гибридные хранилища данных (HOLAP – Hybrid OLAP). Виртуальные хранилища данных. Введение в процесс ETL(Extraction, Transformation, Loading). Извлечение данных в ETL. Преобразование данных в ETL. Загрузка данных в хранилище. Особенности загрузки данных из локальных источников. Обогащение данных.

Тема 3. Трансформация данных.

Введение в трансформацию данных. Особенности трансформации временных рядов. Группировка и разгруппировка данных. Слияние данных. Квантование. Нормализация и кодирование данных.

Тема 4. Визуализация данных.

Введение в визуализацию данных. Визуализаторы общего назначения. OLAP-анализ. Визуализаторы, применяемые для оценки качества моделей. Визуализаторы, применяемые для интерпретации результатов анализа.

Модуль 2. Инструменты Data mining

Тема 5. Оценка качества, очистка и предобработка данных.

Введение в оценку качества данных (ADQ – Assessment Data Quality). Технологии и методы оценки качества данных. Очистка и предобработка данных. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановление пропущенных значений.

Введение в сокращение размерности. Сокращение числа признаков.
Сокращение числа значений признаков и записей. Сэмплинг.

Тема 6. Поиск ассоциативных правил.

Введение в аффинитивный анализ (affinity analysis). Алгоритм a priori.
Иерархические ассоциативные правила.

Тема 7. Кластеризация.

Введение в кластеризацию. Классификация методов кластеризации.
Алгоритм кластеризации k-means. Сети Кохонена (KCN – Kohonen network).
Карты Кохонена (SOM – self organizing map). Проблемы алгоритмов
кластеризации.

Тема 8. Классификация и регрессия. Статистические методы

Введение в классификацию и регрессию. Простая линейная регрессия.
Оценка соответствия простой линейной регрессии реальным данным.
Простая регрессионная модель. Множественная линейная регрессия. Модель
множественной линейной регрессии. Регрессия с категориальными входными
переменными. Методы отбора переменных в регрессионные модели.
Ограничения применимости регрессионных моделей. Основы логистической
регрессии. Интерпретация модели логистической регрессии. Множественная
логистическая регрессия.

Модуль 3. Анализ и прогнозирование временных рядов.

Тема 9. Классификация и регрессия. Машинное обучение

Введение в деревья решений. Алгоритмы построения деревьев
решений. Алгоритмы ID3 и C4.5. Алгоритм CART. Упрощение деревьев
решений. Введение в нейронные сети. Искусственный нейрон. Принципы
построения нейронных сетей. Алгоритмы обучения нейронных сетей.
Алгоритм обратного распространения ошибки.

Тема 10. Анализ и прогнозирование временных рядов.

Временной ряд и его компоненты. Модели прогнозирования.
Прогнозирование в торговле и логистике.

Тема 11. Ансамбли моделей.

Введение в ансамбли моделей. Бэггинг. Бустинг. Альтернативные
методы построения ансамблей.

Тема 12. Сравнение моделей.

Оценка эффективности и сравнение моделей. Оценка ошибки модели.
Издержки ошибочной классификации. Lift- и Profit-кривые. ROC-анализ.

2.3. Темы практических и/или семинарских занятий

Модуль 1. Технологии обеспечения качественного анализа данных

Тема 1. Технология KDD (Knowledge Discovery in Databases).

Тема 2. Многомерные хранилища данных (MOLAP – Multidimensional OLAP). Гибридные хранилища данных (HOLAP – Hybrid OLAP). Виртуальные хранилища данных.

Тема 3. Контрольная работа №1 – Формы представления, типы и виды анализируемых данных. Источники данных для анализа.

Модуль 2. Инструменты Data mining

Тема 1. Введение в технологию Data Mining. Программный инструментарий анализа данных.

Тема 2. Введение в визуализацию данных. Визуализаторы общего назначения. OLAP-анализ. Визуализаторы, применяемые для оценки качества моделей.

Тема 3. Контрольная работа №2 – Технологии и методы оценки качества данных.

Модуль 3. Анализ и прогнозирование временных рядов.

Тема 1. Временной ряд и его компоненты. Модели прогнозирования. Введение в ансамбли моделей.

Тема 2. Контрольная работа №3 – Оценка эффективности и сравнение моделей. Оценка ошибки модели. Издержки ошибочной классификации. Lift- и Profit-кривые. ROC-анализ.

2.4. Лабораторные работы

Лабораторная работа № 1 «Основы работы с аналитической платформой Deductor studio» (2 часа)

Программно-аппаратное обеспечение: ПЭВМ IBM PC (операционная система Windows XP), аналитическая платформа Deductor Studio.

Лабораторная работа № 2 «Трансформация данных в Deductor Studio» (2 часа)

Программно-аппаратное обеспечение: ПЭВМ IBM PC (операционная система Windows XP), аналитическая платформа Deductor Studio.

Лабораторная работа № 3 «Создание, заполнение и использование хранилища данных Deductor Warehouse на базе Firebird» (2 часа)

Программно-аппаратное обеспечение: ПЭВМ IBM PC (операционная система Windows XP), аналитическая платформа Deductor Studio Academic.

Лабораторная работа № 5 «Определение представления источника данных в проекте служб Analysis Services» (2 часа)

Программно-аппаратное обеспечение: ПЭВМ IBM PC (операционная система Windows XP Professional), SQL Server 2008 Developer.

Лабораторная работа № 6 «Определение и развертывание куба» (4 часа)

Место проведения: специализированный компьютерный класс.

Программно-аппаратное обеспечение: ПЭВМ IBM PC (операционная система Windows XP Professional), SQL Server 2008 Developer.

Лабораторная работа № 7 «Изменение мер, атрибутов и иерархий» (4 часа)

Программно-аппаратное обеспечение: ПЭВМ IBM PC (операционная система Windows XP Professional), SQL Server 2008 Developer.

5. Образовательные технологии

С целью формирования и развития профессиональных навыков обучающихся в соответствии с требованиями ФГОС ВО по направлению подготовки предусматривается широкое использование в учебном процессе активных и интерактивных форм проведения занятий:

- во время лекционных занятий используется презентация с применением слайдов с графическим и табличным материалом, что повышает наглядность и информативность используемого теоретического материала;

- практические занятия предусматривают использование групповой формы обучения, которая позволяет студентам эффективно взаимодействовать в микрогруппах при обсуждении теоретического материала;

- использование кейс–метода (проблемно–ориентированного подхода), то есть анализ и обсуждение в микрогруппах конкретной деловой ситуации из практического опыта товароведной деятельности отечественных и зарубежных кампаний;

- использование тестов для контроля знаний во время текущих аттестаций и промежуточной аттестации;

- подготовка рефератов и докладов по самостоятельной работе студентов и выступление с докладом перед аудиторией, что способствует формированию навыков устного выступления по изучаемой теме и активизирует познавательную активность студентов.

Предусмотрены также встречи с представителями предпринимательских структур, государственных и общественных организаций, мастер-классы специалистов.

6. Учебно-методическое обеспечение самостоятельной работы студентов.

Возрастает значимость самостоятельной работы студентов в межсессионный период. Поэтому изучение курса «Анализ данных» предусматривает работу с основной специальной литературой, дополнительной обзорного характера, а также выполнение домашних заданий.

Самостоятельная работа студентов должна способствовать более глубокому усвоению изучаемого курса, формировать навыки исследовательской работы и ориентировать студентов на умение применять теоретические знания на практике.

Задания для самостоятельной работы, их содержание и форма контроля приведены в форме таблицы.

Наименование тем	Содержание самостоятельной работы	Форма контроля
Тема 1.1. Технологии анализа данных.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата, проверка конспекта
Тема 1.2. Консолидация данных.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата.
Тема 1.3. Трансформация данных.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата.
Тема 1.4. Визуализация данных.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата.
Тема 2.1. Оценка качества, очистка и предобработка данных.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата.
Тема 2.2. Поиск ассоциативных правил.	Работа с учебной литературой. Подготовка реферата. Решение задач и тестов	Опрос, оценка выступлений, защита реферата. Проверка заданий.

Тема 2.3. Кластеризация. Тема 2.4. Классификация и регрессия. Статистические методы.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата. Проверка конспекта.
Тема 2.5. Оценка качества, очистка и предобработка данных.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата. Проверка конспекта.
Тема 3.1. Классификация и регрессия. Машинное обучение.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата.
Тема 3.2.. Анализ и прогнозирование временных рядов.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата.
Тема 3.3. Ансамбли моделей.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата.
Тема 3.4. Сравнение моделей.	Работа с учебной литературой. Подготовка реферата.	Опрос, оценка выступлений, защита реферата.

Целью подготовки реферата является приобретение навыков творческого обобщения и анализа имеющейся литературы по рассматриваемым вопросам, что обычно является первым этапом самостоятельной работы. По каждому модулю предусмотрены написание и защита одного реферата. Всего по дисциплине студент может представить шесть рефератов. Тему реферата студент выбирает самостоятельно из предложенной тематики. При написании реферата надо составить краткий план, с указанием основных вопросов избранной темы. Реферат должен включать введение, несколько вопросов, посвященных рассмотрению темы, заключение и список использованной литературы. В вводной части реферата следует указать основания, послужившие причиной выбора данной темы, отметить актуальность рассматриваемых в реферате вопросов. В основном разделе излагаются наиболее существенные сведения по теме, производится их анализ, отмечаются отдельные недостатки или нерешенные еще вопросы, вносятся и обосновываются предложения по повышению качества потребительских товаров, расширению ассортимента, совершенствованию

контроля за качеством и т.д. В заключении реферата на основании изучения литературных источников должны быть сформулированы краткие выводы и предложения. Список литературы оформляется в соответствии с требованиями ГОСТ 7.1-84 «Библиографическое описание документа». Перечень литературы составляется в алфавитном порядке фамилий первых авторов, со сквозной нумерацией. Примерный объем реферата 15-20 страниц.

Предусмотрено проведение индивидуальной работы (консультаций) со студентами в ходе изучения материала данной дисциплины.

Тематика рефератов:

1. Основные положения концепции хранилищ данных.
2. Многомерное представление данных. Правила Кодда, определяющие OLAP-системы.
3. Практическое применение алгоритмов Data mining.
4. Классификация с несколькими независимыми переменными методом Naïve Bayes.
5. Поиск оптимальной функции методом наименьших квадратов.
6. Сиквенциальный анализ.
7. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации.
8. Кластеризация данных при помощи нечетких отношений.
9. Стандарты Data mining. Характеристика стандартов CWM и PMLL.
10. Библиотеки доступа к алгоритмам Data mining. Характеристика библиотеки Xelopes.
11. Характеристика программных инструментов для выполнения интеллектуального анализа данных.
12. Характеристика OLTP-систем.
13. Предпосылки появления и разновидности систем поддержки принятия решений.
14. Характеристика реляционных хранилищ данных.
15. Характеристика многомерных хранилищ данных.
16. Характеристика гибридных хранилищ данных.
17. Характеристика виртуальных хранилищ данных.
18. Технологии и методы оценки качества данных. Профайлинг.

19. Типичный набор инструментов преобработки данных в аналитическом приложении.
20. Алгоритмы и методы сокращения числа признаков.
21. Общая характеристика и классификация методов кластерного анализа данных.
22. Анализ данных с использованием методов классификации и регрессии.
23. Цели, задачи и принципы построения деревьев решений. Общая характеристика алгоритмов построения деревьев решений.
24. Сферы применения деревьев решений.
25. Цели, задачи и принципы работы нейронных сетей.
26. Алгоритмы обучения нейронных сетей.
27. Цели, задачи, принципы и модели прогнозирования.
28. Трансформация данных на разных этапах аналитического процесса. Типичные средства трансформации.
29. Характеристика классов задач, решаемых методами Data Mining.
30. Характеристика технологии Knowledge Discovery in Databases (KDD).

3.2.Перечень вопросов для самостоятельного изучения

1. Основные методы оценивания параметров распределения случайной величины.
2. Несмещенная, состоятельная, эффективная оценки.
3. Статистические гипотезы.
4. Порядок проверки параметрических статистических гипотез?
5. Однородные выборки.
6. Основные типы неоднородности выборок.
7. Шкалы измерения и допустимые преобразования в этих шкалах.
8. Таблица сопряженности признаков.
9. Независимость признаков, измеряемых в номинальной шкале.

10. Определения и основные свойства коэффициента корреляции.

11. Задача линейной регрессии.

12. Метод наименьших квадратов (МНК).

13. Методы оценивания параметров регрессии.

7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.

7.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.

Перечень компетенций с указанием этапов их формирования приведен в описании образовательной программы.

Компетенция	Знания, умения, навыки	Процедура освоения
ОК-16	Знать: основные принципы сбора и подготовки исходных данных; Уметь: систематизировать и обобщать информацию; Владеть: основными приемами сбора информации и инструментальными средствами для обработки информации.	Устный опрос, написание рефератов, тестирование
ПК-19	Знать: основные методы анализа данных; Уметь: применять современные математико-статистические методы и инструментальные средства для исследования объектов профессиональной деятельности; Владеть: методами статистического анализа и прогнозирования.	Устный опрос, написание рефератов, тестирование
ПК-20	Знать: основные методы анализа данных; Уметь: использовать математические, инструментальные средства для обработки, анализа и систематизации информации по теме исследования;	Устный опрос, проведение деловой игры, написание рефератов, тестирование

	Владеть: навыками проведения факторного и кластерного анализа.	
--	---	--

7.2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания.

ОК-16 (способен работать с информацией из различных источников).

Уровень	Показатели (что обучающийся должен продемонстрировать)	Оценочная шкала		
		Удовлетворительно	Хорошо	Отлично
Пороговый	<p>Знать: основные принципы сбора и подготовки исходных данных;</p> <p>Уметь: систематизировать и обобщать информацию;</p> <p>Владеть: основными приемами сбора информации и инструментальными средствами для обработки информации.</p>	<p>Имеет неполное представление об основных принципах сбора и подготовки исходных данных</p> <p>Демонстрирует слабое умение систематизировать и обобщать информацию.</p> <p>Слабо владеет основными приемами сбора информации и инструментальными средствами для обработки информации.</p>	<p>Допускает неточности в понимании и представлении об основных принципах сбора и подготовки исходных данных</p> <p>Может систематизировать и обобщать информацию информационных систем и технологий</p> <p>Владеет основными приемами сбора информации и инструментальными средствами для обработки информации.</p>	<p>Демонстрирует четкое представление основных принципов сбора и подготовки исходных данных</p> <p>Может правильно анализировать и систематизировать и обобщать информацию</p> <p>Эффективно владеет основными приемами сбора информации и инструментальными средствами для обработки информации.</p>

ПК-19 (использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования).

Уровень	Показатели (что обучающийся должен продемонстрировать)	Оценочная шкала		
		Удовлетворительно	Хорошо	Отлично
Пороговый	<p>Знать: основные методы анализа данных;</p> <p>Уметь: применять в современные математико-статистические методы и инструментальные средства для исследования объектов профессиональной деятельности;</p> <p>Владеть: методами статистического анализа и прогнозирования.</p>	<p>Имеет неполное представление об основных методах анализа данных.</p> <p>Демонстрирует слабое умение применять современные математико-статистические методы и инструментальные средства для исследования объектов профессиональной деятельности</p> <p>Слабо владеет методами статистического анализа и прогнозирования.</p>	<p>Допускает неточности в знании основных методов анализа данных.</p> <p>Может анализировать и применять современные математико-статистические методы и инструментальные средства для исследования объектов профессиональной деятельности.</p> <p>Владеет методами статистического анализа и прогнозирования.</p>	<p>Демонстрирует четкое представление об основных методах анализа данных.</p> <p>Может грамотно анализировать и применять современные математико-статистические методы и инструментальные средства для исследования объектов профессиональной деятельности</p> <p>Эффективно владеет методами статистического анализа и прогнозирования.</p>

ПК-20 (использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования).

Уровень	Показатели (что обучающийся)	Оценочная шкала		
		Удовлетворительно	Хорошо	Отлично

	должен продемонстрировать)	бно		
Пороговый	<p>Знать: основные методы анализа данных.</p> <p>Уметь: использовать математические, инструментальные средства для обработки, анализа и систематизации информации по теме исследования.</p> <p>Владеть: навыками проведения факторного кластерного анализа.</p>	<p>Имеет неполное представление об основных методах анализа данных.</p> <p>Демонстрирует слабое умение применять математические, инструментальные средства для обработки, анализа и систематизации информации по теме исследования.</p> <p>Слабо владеет навыками проведения факторного кластерного анализа.</p>	<p>Допускает неточности в понимании основных методов анализа данных.</p> <p>Может применять математические, инструментальные средства для обработки, анализа и систематизации информации по теме исследования.</p> <p>Владеет методиками и навыками проведения факторного кластерного анализа.</p>	<p>Демонстрирует четкое представление об основных методах анализа данных.</p> <p>Может грамотно использовать математические, инструментальные средства для обработки, анализа и систематизации информации по теме исследования.</p> <p>Эффективно владеет навыками проведения факторного кластерного анализа.</p>

Если хотя бы одна из компетенций не сформирована, то положительная оценки по дисциплине быть не может.

7.3. Типовые контрольные задания

Текущий контроль успеваемости в форме опросов, рефератов, дискуссий, тестов, решения задач и промежуточный контроль в форме экзамена.

Контрольные вопросы к зачету для промежуточного контроля

1. Модели и их свойства. Аналитический и информационный подходы к моделированию.

2. Формы представления, типы и виды анализируемых данных.
3. Обучение моделей «с учителем» и «без учителя». Обучающее и тестовое множество. Ошибки обучения. Эффект переобучения.
4. Общая схема анализа данных. Требования к алгоритмам анализа данных.
5. Основные принципы сбора (формализации) данных. Требования к объемам анализируемых данных.
6. Характеристика этапов технологии KDD.
7. Data Mining. Характеристика классов задач, решаемых методами Data Mining.
8. Программный инструментарий для выполнения анализа данных.
9. Цели, задачи и основное содержание консолидации данных. Обобщенная схема процесса консолидации.
10. Характеристика OLTP-систем.
11. Предпосылки появления систем поддержки принятия решений DSS. Понятие ESS, EIS и GDSS.
12. Основные положения концепции хранилищ данных (DW).
13. Реляционные хранилища данных (ROLAP).
14. Технология OLAP. Сущность многомерного представления данных.
15. Структура многомерного куба. Работа с измерениями.
16. Многомерные хранилища данных (MOLAP).
17. Гибридные хранилища данных (HOLAP).
18. Виртуальные хранилища данных.
19. Цели, задачи и основное содержание процесса ETL.
20. Основные виды проблем в данных, из-за которых они нуждаются в очистке.
21. Организация процесса загрузки данных в хранилище. Постзагрузочные операции.
22. Причины отказа от использования хранилищ данных. Особенности загрузки данных из локальных источников.
23. Обогащение данных.

24. Цели, задачи и основное содержание трансформации данных. Трансформация данных на разных этапах аналитического процесса. Типичные средства трансформации.
25. Особенности трансформации временных рядов. Скользящее окно. Преобразование даты и времени.
26. Группировка и разгруппировка данных.
27. Способы слияния данных.
28. Квантование данных.
29. Нормализация и кодирование данных.
30. Цели, задачи и основное содержание визуализации данных. Группы методов визуализации.
31. Визуализаторы общего назначения. OLAP-анализ.
32. Манипуляции с измерениями OLAP-куба.
33. Визуализаторы, применяемые для оценки качества моделей.
34. Визуализаторы, применяемые для интерпретации результатов анализа.
35. Технологии и методы оценки качества данных. Профайлинг.
36. Очистка и предобработка данных.
37. Типичный набор инструментов предобработки данных в аналитическом приложении.
38. Фильтрация данных. Обработка дубликатов и противоречий.
39. Выявление аномальных и восстановление пропущенных значений.
40. Алгоритмы и методы сокращения числа признаков.

Задачи по дисциплине

1. Выборка имеет распределение . Найдите оценку максимального правдоподобия параметра . Докажите несмещённость и состоятельность построенной оценки.
2. Стоимость лечения одного пациента-льготника с диагнозом «доуденит» в областях Дальневосточного

федерального округа (в тыс.руб. на ноябрь 2013 года)
составила:

Амурская область	245,61
Еврейская АО	101,45
Камчатская область	202,84
Корякская АО	327,63
Магаданская область	144,5
Приморский край	458,81

Предполагая, что наблюдения имеют гауссовское распределение, постройте доверительный интервал уровня надежности 0,95 для средней стоимости лечения в Дальневосточном ФО пациента с указанным диагнозом.

3. В результате проведенного исследования было установлено, что у 309 светлоглазых мужчин жены так же имеют светлые глаза, а у 214 светлоглазых мужчин жены темноглазые. У 119 темноглазых мужчин жены так же темноглазые, а у 132 темноглазых мужчин жены светлоглазые. Имеется ли зависимость между цветом глаз мужей и жен?

4. В таблице представлены данные за 1995 год показателей X (ВВП в паритетах покупательной способности) и Y(коэффициент детской смертности в %) для следующих стран: Бурунди, Чад, Индия, Египет, Мексика, Бразилия, Республика Корея, Канада, США, Швейцария.

X	2.3	2.6	5.2	12.2	23.7	20	42.2	78.3	100	95.9
Y	98	117	68	16	33	44	10	6	8	6

Считая, что наблюдения имеют гауссовское распределение, выясните являются ли признаки X и Y зависимыми.

5. Анализируется прибыль Y (млн. \$) в зависимости от расходов X (млн. \$) на рекламу. Данные наблюдений за 4 года представлены в таблице:

X	0,8	2,5	4,0	5,7
Y	5	15	20	25

Оцените по методу наименьших квадратов коэффициенты линейной регрессии вида $Y =$

7.4. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Оценка за модуль определяется как сумма баллов за текущую и контрольную работу.

Коэффициент весомости баллов, набранных за текущую и контрольную работу, составляет 0,5/0,5.

Текущая работа включает оценку аудиторной и самостоятельной работы.

Оценка знаний студента на практическом занятии (аудиторная работа) производится по 100-балльной шкале.

Оценка самостоятельной работы студента (написание эссе, подготовка доклада, выполнение домашней контрольной работы и др.) также осуществляется по 100-балльной шкале.

Для определения среднего балла за текущую работу суммируются баллы, полученные за аудиторную и самостоятельную работу, полученная сумма делится на количество полученных оценок.

Итоговый балл за текущую работу определяется как произведение среднего балла за текущую работу и коэффициента весомости.

Если студент пропустил занятие без уважительной причины, то это занятие оценивается в 0 баллов и учитывается при подсчете среднего балла за текущую работу.

Если студент пропустил занятие по уважительной причине, подтвержденной документально, то преподаватель может принять у него отработку и поставить определенное количество баллов за занятие. Если преподаватель по тем или иным причинам не принимает отработку, то это занятие при делении суммарного балла не учитывается.

Контрольная работа за модуль также оценивается по 100-балльной шкале. Итоговый балл за контрольную работу определяется как произведение баллов за контрольную работу и коэффициента весомости.

Критерии оценок аудиторной работы студентов по 100-балльной шкале:

«0 баллов» - студент не смог ответить ни на один из поставленных вопросов

«10-50 баллов» - обнаружено незнание большей части изучаемого материала, есть слабые знания по некоторым аспектам рассматриваемых вопросов

«51-65 баллов» - неполно раскрыто содержание материала, студент дает ответы на некоторые рассматриваемые вопросы, показывает общее понимание, но допускает ошибки

«66-85 баллов» - студент дает почти полные ответы на поставленные вопросы с небольшими проблемами в изложении. Делает самостоятельные выводы, имеет собственные суждения.

«86-90 баллов» - студент полно раскрыл содержание материала, на все поставленные вопросы готов дать абсолютно полные ответы, дополненные собственными суждениями, выводами. Студент подготовил и отвечает дополнительный материал по рассматриваемым вопросам.

Таблица перевода рейтингового балла в «5»-балльную шкалу

Итоговая сумма баллов по дисциплине по 100-балльной шкале	Оценка по 5-балльной шкале
0-50	Неудовлетворительно
51-65	Удовлетворительно
66-85	Хорошо
86-100	Отлично

Таблица перевода рейтингового балла по дисциплине в «зачтено»

или «не зачтено»

Итоговая сумма баллов по дисциплине по 100-балльной шкале	Оценка по дисциплине
0-50	Не зачтено
51-100	Зачтено

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.

1.7. Учебно-методическое и информационное обеспечение дисциплины.

а) основная литература

1. Паклин Н.Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям. – СПб.: Питер, 2009
2. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. 2-е изд.– СПб.: БХВ – Петербург, 2008
3. Кацко И.А., Н.Б. Паклин. Практикум по анализу данных на компьютере. – М.: КолосС, 2009

б) дополнительная литература:

4. Дюк В.А., Самойленко А.П. Data Mining: учебный курс. – СПб.: Питер, 2001
5. Хайкин С. Нейронные сети: полный курс. 2-е. изд. / Пер. с англ. – М.: Издательский дом «Вильямс», 2006.
6. Ханк Д.Э., Уичерн Д.У., Райте А.Дж. Бизнес-прогнозирование. 7-е изд. / Пер. с англ. – М.: Издательский дом «Вильямс», 2003
7. Дубров А. М., Мхитарян В. С., Трошин Л. И. Многомерные статистические методы: Учебник. – М.: Финансы и статистика, 2000
8. Нейронные сети. STATISTICA Neural Network: Методология и технологии современного анализа данных/ под ред. В.П.Боровикова. – М.: Горячая линия-Телеком, 2008
9. Халафян А.А. STATISTICA 6. Статистический анализ данных: Учебник - 3-е изд. – М.: Бином-Пресс, 2007
10. Андрейчиков А.В., Андрейчикова О.Н. Интеллектуальные информационные системы. – М.: ФиС, 2004
11. Кравченко Т.К., Перминов Г.И. Информационная технология процесса принятия экономических решений. - М.: ГУ-ВШЭ., 2005
12. Наследов А.Д. SPSS: Компьютерный анализ данных в психологии и социальных науках - 2-е изд. – СПб.: Питер, 2007

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

1. Федеральный портал //
2. Электронные каталоги Научной библиотеки ДГУ // //

10. Методические указания для обучающихся по освоению дисциплины.

Для успешного освоения учебного материала курса «Анализ данных» требуются систематическая работа по изучению лекций и рекомендуемой литературы, решению домашних задач и домашних контрольных работ, а также активное участие в работе семинаров.

Показателем освоения материала служит успешное решение задач предлагаемых домашних контрольных работ и выполнение аудиторных самостоятельных и контрольных работ. В качестве оценочных средств программой дисциплины предусматривается:

текущий контроль (аудиторные контрольные работы, домашние задания).

промежуточный контроль (зачет).

Промежуточный контроль изучения дисциплины «Анализ данных» проводится в форме письменного зачет в 4-м семестре. Итоговая оценка за зачет выставляется в форме «зачет» и «незачет» в баллах по 100-балльной шкале:

«незачет» - менее 51 балла;

«зачет» - от 51 до 100 баллов.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.

При осуществлении образовательного процесса рекомендуются компьютерные технологии, основанные на операционных системах прикладные программы а также сайты образовательных учреждений и журналов, информационно-справочные системы, электронные учебники.

При проведении занятий рекомендуется использовать компьютеры, мультимедийные проекторы, интерактивные экраны.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

Университет обладает достаточной базой аудиторий для проведения всех занятий, предусмотренных образовательной программой дисциплины математический анализ. Кроме того, на экономических факультетах имеются компьютерные и учебные классы, оснащенные компьютерами с соответствующим программным обеспечением и мультимедив – проекторами.

В университете имеется необходимый комплект лицензионного программного обеспечения.