

---

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**

Государственное бюджетное образовательное учреждение высшего профессионального образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ»  
(ТУСУР)

УТВЕРЖДАЮ

Заведующий кафедрой ЭМИС

\_\_\_\_\_ И. Г. Боровской

« \_\_\_ » \_\_\_\_\_ 2012 г.

**С.И. КОЛЕСНИКОВА**

**АНАЛИЗ ДАННЫХ**

***Методические указания по самостоятельной работе студентов***

2012

Составитель: Колесникова С.И., каф.ЭМИС

## АННОТАЦИЯ

Цели настоящих методических указаний: 1) освоение основных понятий и определений раздела знаний «Анализ данных»; 2) приобретение практических навыков в построении алгоритмов анализа данных, поиска закономерностей и распознавания характерных образов, анализа качества алгоритмов. В четырех частях указаний приведены примеры задач и методов их решения (анализа возможного решения) на следующие темы:

1. Методы и средства анализа данных.
2. Методы классификации и кластеризации.
3. Методы построения математических моделей и прогнозирования временных рядов.
4. Основные модели управления данными, многомерный анализ данных.

Теоретический материал приведен только тот и в том объеме, который необходим для решения предлагаемых задач. Задачи контрольных заданий являются весьма простыми, они предназначены для усвоения основных начальных понятий и основ современных методов анализа данных. Предполагается, что студенты знают математику в объеме, требуемом в техническом ВУЗе.

Методические указания предназначены для студентов экономического факультета.

**СОДЕРЖАНИЕ МЕТОДИЧЕСКИХ УКАЗАНИЙ  
К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ  
по дисциплине «Анализ данных»**

**ОГЛАВЛЕНИЕ**

1. УКАЗАНИЯ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ по разделу 1 .....	5
1.1. Выполнение индивидуальных домашних заданий (ИДЗ) №1 по теме: «Методы и средства анализа данных».....	5
1.2. Типовые тесты к разделу 1 (пример) .....	5
1.3. Подготовка к интерактивному занятию №1 «Анализ реальной проблемы и выбор методов и средства анализа данных. Основы непараметрической статистики» .....	6
2. УКАЗАНИЯ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ по разделу 2 .....	7
2.1. Выполнение индивидуальных домашних заданий (ИДЗ) №2 по теме: «Методы классификации и кластеризации».....	7
2.2. Типовые тесты к разделу 2 (пример) .....	8
2.3. Подготовка к интерактивному занятию №2 «Классификация на базе конструирования нейросети в специализированном нейрорпакете» (8 ч).....	10
3. УКАЗАНИЯ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ по разделу 3 .....	11
3.1. Выполнение индивидуальных домашних заданий (ИДЗ) №3 по теме: «Методы построения математических моделей и прогнозирования временных рядов. Анализ свойств одномерных хаотических моделей».....	11
3.2. Типовые тесты к разделу 3 (пример) .....	12
3.3. Подготовка к интерактивному занятию №3 «Методы анализа нестационарных стохастических временных рядов».....	13
4. УКАЗАНИЯ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ по разделу 4 .....	14
4.1. Выполнение индивидуальных домашних заданий (ИДЗ) №4 по теме: «Основные модели управления данными, многомерный анализ данных» .....	14
4.2. Типовые тесты к разделу 4 (пример) .....	14
4.3. Подготовка к интерактивному занятию №4 «Основные модели управления данными и распределённый анализ данных».....	16
Использованная литература .....	18

№ п/п	№ раздела дисциплины	Вид самостоятельной работы	Трудоемкость (час.)	Компетенции ОК, ПК	Контроль выполнения работы (Опрос, тест, Индивидуальные)
-------	----------------------	----------------------------	---------------------	--------------------	--

Составитель: Колесникова С.И., каф.ЭМИС

	из табл . 5.1				домашние задания (ИДЗ), и т.д)
1.	1	Подготовка к проверочной индивидуальной работе и ее выполнение	12	ОК-12, ОК-13, ПК-14, ПК-15	Отчет по ИДЗ. Отчет по ИнЗ.
2	1	Подготовка к интерактивному занятию №1	8	ОК-12, ОК-13, ПК-14, ПК-15	
	2	Подготовка к проверочной индивидуальной работе и ее выполнение	12	ОК-12, ОК-13, ПК-14, ПК-15	Отчет по ИДЗ. Отчет по ИнЗ.
3	2	Подготовка к интерактивному занятию №2	8	ОК-12, ОК-13, ПК-14, ПК-15	
4	3	Подготовка к проверочной индивидуальной работе и ее выполнение	12	ОК-12, ОК-13, ПК-14, ПК-15	Отчет по ИДЗ. Отчет по ИнЗ.
5	3	Подготовка к интерактивному занятию №3	8	ОК-12, ОК-13, ПК-14, ПК-15	
6	4	Подготовка к проверочной индивидуальной работе и ее выполнение	18	ОК-12, ОК-13, ПК-14, ПК-15	Отчет по ИДЗ. Отчет по ИнЗ..
7	4	Подготовка к интерактивному занятию №4	12	ОК-12, ОК-13, ПК-14, ПК-15	
8	1-4	Подготовка к тестированию и экзамену	36	ОК-12, ОК-13, ПК-14, ПК-15	

**Обозначения:** ИДЗ - индивидуальные домашние задания  
СРС - самостоятельная работа студентов  
ИнЗ - интерактивное занятие

З-Эл – знания элементарные (определения, понятия, умение приводить иллюстрирующие примеры);

З-Пр – знания продуктивные (умение применить знания элементарные для решения учебных задач);

У-Эл – «умения» элементарные (уметь пользоваться готовыми частными алгоритмами для решения типовых задач), умение решать задачи по шаблону (копировать);

У-Пр – «умения» продуктивные (применять положения и известные частные алгоритмы дисциплины для решения практических задач);

В-Эл – элементарное владение методами дисциплины и уверенное осуществление (построение) основных операций для решения типовых задач;

В-Пр – продуктивно распознавать проблемы, алгоритмизировать их анализ и применять методы дисциплины для решения практических задач;

**Замечание.** В методических указаниях использовался материал вышеуказанной основной и дополнительной литературы рабочей программы, а также материал других источников и интернет-ресурсов:

## 1. УКАЗАНИЯ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ по разделу 1

### 1.1. Выполнение индивидуальных домашних заданий (ИДЗ) №1 по теме: «Методы и средства анализа данных»

*Цель занятия:* Проведение исследований на базе поискового метода по одной из ниже заданных тем. Знакомство с основными математическими дисциплинами, изучающими закономерности массовых случайных явлений и основными этапами исследования любой сложной (плохоформализуемой) проблемы:

- 1) определение проблемы;  
разработка подхода (подходов в случае коллективного принятия решения) к решению проблемы;
- 2) разработка плана исследования;
- 3) сбор (определение) данных (обучающих и контрольных)
- 4) предобработка данных;
- 5) исследование и решение задачи;
- 6) подготовка отчета и презентации-доклада.

*Форма текущего контроля* освоения компетенций ОК-12, ОК-13 (уровни З-Эл; У-Эл; В-Эл), ПК-14, ПК-15 (уровни З-Пр, У-Пр, В-Пр): *отчет* по выбранной теме и подготовка презентации-защиты:

Рекомендуемые темы:

- Постановка задачи анализа данных. Сущность и условия применимости методов анализа данных.
- Модели Data Mining.
- Классификация методов анализа данных.
- Системы распознавания образов на базе Data Mining и принципы их построения.
- Примеры практического применения Data Mining.
- Статистические методы оценивания экономических показателей.
- Методы Data Mining в экономике.
- Методы Data Mining в банковской деятельности.
- Виды СППР на базе Data Mining.

### 1.2. Типовые тесты к разделу 1 (пример)

№	Вопросы	Ответы
1.	Термин «анализ данных» означает 1. Извлечение данных из многих разнородных источников, представленных в различных форматах, приведение к единому формату и структуре. 2. Организация хранения и предоставления необходимых пользователям сведений. 3. Собственно анализ состоит из оперативного и интеллектуального, а также в формировании типовых документов. 4. Подготовка результатов анализа всех видов для эффективного восприятия потребителями.	4
2.	Информационное пространство - это: 1. набор сведений о системе или объекте. 2. совокупность информационных объектов, информационно отображающих свойства системы и протекающие в ней процессы.	2

Составитель: Колесникова С.И., каф.ЭМИС

3.	Информационное пространство состоит из следующих единиц информации: 1. Бит, байт 2. реквизит. 3. показатель 4. составная единица информации 5. база данных	2-5
4.	Этапы прохождения данных из источников – это:  <b>Ответ:</b> Первый - извлечения, преобразования и загрузки данных. Второй - накопления, обеспечения готовности данных к использованию. Третий - применения данных, содержащихся в хранилище и извлекаемых напрямую из первичных источников.	
5.	Идея гибкой архитектуры данных означает, что: 1. архитектура данных в информационно-аналитической системе может быть легко изменена 2. любому пользователю из числа доверенных лиц должна быть обеспечена возможность доступа к любому разрешённому для использования участку данных, которыми располагает предприятие (организация)	2
6.	Основные принципы построения информационных хранилищ – следующие: 1. предметная ориентированность 2. многомерность 3. интегрированность 4. неизменчивость 5. поддержка хронологии Найдите не относящийся к информационным хранилищам признак.	
7.	Многомерные схемы данных информационно-аналитической системе бывают следующих видов: 1. схема “звезда” 2. схема "снежинка" 3. схема "капля" 4. схема "созвездие"	1, 2
8.	<b>Обучение без учителя (самообучение)</b>  1) процесс формирования обобщенных образов классов, на основе обучающей выборки, содержащей характеристики конкретных объектов как в описательных, так и в классификационных шкалах и градациях. 2) процесс формирования обобщенных образов классов, на основе обучающей выборки, содержащей характеристики конкретных объектов, причем только в описательных шкалах и градациях. 3) этот процесс называют кластерным анализом (таксономией).	

### 1.3. Подготовка к интерактивному занятию №1 «Анализ реальной проблемы и выбор методов и средства анализа данных. Основы непараметрической статистики»

**Цель занятия:** активное воспроизведение ранее полученных знаний по разделу 1 «Анализ реальной проблемы и выбор методов и средства анализа данных. Основы непараметрической статистики» в «незнакомых» условиях (применение основных понятий темы раздела 1 для решения задачи: построение моделей для анализа данных в практических задачах).

**Форма текущего контроля** освоения компетенций (ОК-1, ОК-3, ОК-12, уровни З-Эл, У-Эл, В-Эл; ПК-26 уровни З-Пр, У-Пр, В-Пр) (см. табл.1 методических указаний к практическим занятиям): *отчет* по решению практических текстовых задач.

**Ознакомиться со следующим материалом (по указанным источникам):**

- 1) Схема исследования сложных динамических объектов с вероятностно-статистической точки зрения;
- 2) Методы факторного анализа.
- 3) Метод главных компонент.

**Подготовить отчет команды, сформированной на ИнЗ 1, по обсуждаемым задачам, содержащий положения:**

- 1) Постановка решаемых задач.
- 2) Изложение обзора вариантов их решения.
- 3) Защита выбранного варианта и решение задач.
- 4) Подготовка презентации-защиты работы команды.

**Ознакомиться с формами текущего контроля**

Таблица 1

№ п / п	Наименование разделов	Формы контроля				
		Знаний	Умений	Навыков	Оценка личностных качеств	Компетенции/ ожидаемый уровень освоения
1	Основы теории вероятностей. Случайные события	Сдача индивидуальных заданий		Отчет по решению реальных практических задач на интерактивном занятии	Соблюдение установленных сроков для отчета и теста	ОК-1, ОК-3, ОК-12/ З-Эл, У-Эл, В-Эл ПК-26/ З-Пр, У-Пр, В-Пр
		Контрольная работа				
		Отчет по ИнЗ №1				

## 2. УКАЗАНИЯ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ по разделу 2

### 2.1. Выполнение индивидуальных домашних заданий (ИДЗ) №2 по теме: «Методы классификации и кластеризации»

**Цель занятия:** Проведение исследований на базе поискового метода по одной из заданных тем.

**Форма текущего контроля** освоения компетенций (ОК-1, ОК-3, ОК-12, уровни З-Эл, У-Эл, В-Эл; ПК-26 уровни З-Пр, У-Пр, В-Пр): *отчет* по выбранной теме и подготовка презентации-защиты:

Рекомендуемые темы:

- Классификация с помощью деревьев решений;
- Байесовская классификация;
- Классификация при помощи искусственных нейронных сетей;

Составитель: Колесникова С.И., каф.ЭМИС

- Классификация методом опорных векторов;
- Статистические методы, в частности, линейная регрессия;
- Классификация при помощи метода ближайшего соседа;
- Классификация сbr-методом (case based reasoning, метод рассуждения по аналогии);
- Классификация при помощи генетических алгоритмов.

## 2.2. Типовые тесты к разделу 2 (пример)

1.	<p><b>Кластеризация –</b></p> <p>1) это операция автоматической классификации, в ходе которой объекты объединяются в группы (кластеры) таким образом, что внутри групп различия между объектами минимальны, а между группами – максимальны. При этом в ходе кластеризации не только определяется состав кластеров, но и сам их набор и границы.</p> <p>2) это операция, основанная на трудно формализуемых знаниях и интуиции исследователя. При этом исследователь сам определяет, какую информацию и каким образом система должна использовать для достижения требуемого эффекта классификации.</p> <p>3) процесс формирования обобщенных образов классов, на основе обучающей выборки, содержащей характеристики конкретных объектов как в описательных, так и в классификационных шкалах и градациях.</p>	
2.	<p><b>Качество решающих правил оценивается</b></p> <p>1) по репрезентативной выборке достаточно полно представляет генеральную совокупность (гипотетическое множество всех возможных объектов каждого образа);</p> <p>2) по обучающей выборке - множеству объектов, заданных значениями признаков и принадлежность которых к тому или иному классу достоверно известна "учителю" и сообщается учителем "обучаемой" системе;</p> <p>3) по контрольной (экзаменационной) выборке, в которую входят объекты, заданные значениями признаков, и принадлежность которых тому или иному образу известна только учителю.</p>	
3.	<p><b>Кластерный анализ (самообучение, обучение без учителя, таксономия) применяется</b></p> <p>1) при автоматическом формировании перечня образов по обучающей выборке;</p> <p>2) в методе потенциальных функций;</p> <p>3) при структурном (лингвистическом) подходе.</p>	
4.	<p><b>Геометрическая интерпретация гипотезы компактности состоит в следующем</b></p> <p>1) объекты, относящиеся к одному классу, расположены «ближе» друг к другу по сравнению с объектами, относящимися к разным классам (таксонам);</p> <p>2) объекты, относящиеся к одному классу, расположены «в среднем ближе» друг к другу по сравнению с объектами, относящимися к разным классам (таксонам);</p>	

	3) объекты, относящиеся к разным классам, можно разделить посредством линейного решающего правила.	
9.	<p><b>Метрика измерения степени близости (расстояния)</b></p> <p>1) должна иметь вид</p> $d_{ij} = \sum_{k=1}^m  x_{ik} - x_{jk} , i, j=1, \dots, n.$ <p>2) должна иметь вид</p> $d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}, i, j=1, \dots, n.$ <p>3) может быть разной, но должна удовлетворять условиям:  <math>d(a, b) = d(b, a)</math>;  <math>d(a, c) \leq d(a, b) + d(b, c)</math>;  <math>d(a, b) \geq 0</math>;  <math>d(a, b) = 0</math> только при <math>a = b</math>.</p>	
10.	<p><b>Метрика измерения степени близости (расстояния)</b></p> <p>1) должна удовлетворять условиям:  <math>d(a, b) = d(b, a)</math>;  <math>d(a, c) \geq d(a, b) + d(b, c)</math>;  <math>d(a, b) \geq 0</math>;  <math>d(a, b) = 0</math> только при <math>a = b</math>.</p> <p>2) должна иметь вид</p> $d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}, i, j=1, \dots, n.$ <p>3) должна удовлетворять условиям:  <math>d(a, b) = d(b, a)</math>;  <math>d(a, c) \leq d(a, b) + d(b, c)</math>;  <math>d(a, b) \geq 0</math>;  <math>d(a, b) = 0</math> только при <math>a = b</math>.</p>	
11.	<p><b>В номинальных измерительных шкалах</b></p> <p>1) отсутствуют отношения порядка, начало отсчета и единица измерения.</p> <p>2) определены отношения "больше – меньше", но отсутствуют начало отсчета и единица измерения.</p> <p>3) определены отношения порядка, все арифметические операции, есть начало отсчета и единица измерения.</p>	
12.	<p><b>На порядковых измерительных шкалах</b></p> <p>1) отсутствуют отношения порядка, начало отсчета и единица измерения.</p> <p>2) определены любые отношения, но отсутствуют начало</p>	

	<p>отсчета и единица измерения.</p> <p>3) определены отношения порядка, все арифметические операции, есть начало отсчета и единица измерения.</p> <p>4) другой ответ.</p>	
13.	<p><b>На шкалах измерительных отношений</b></p> <p>1) отсутствуют отношения порядка, начало отсчета и единица измерения.</p> <p>2) На порядковых шкалах определены отношения "больше – меньше", но отсутствуют начало отсчета и единица измерения.</p> <p>3) На шкалах отношений определены отношения порядка, все арифметические операции, есть начало отсчета и единица измерения.</p>	
14.	<p><b>Применение евклидовой меры расстояний корректно,</b></p> <p>1) если одновременно выполняются два условия:</p> <p>а). все оси координат фазового пространства являются шкалами отношений;</p> <p>б). все оси координат взаимно-перпендикулярны или очень близки к этому;</p> <p>2) оси координат фазового пространства не обязаны быть взаимно-перпендикулярными шкалами отношений;</p> <p>3) оси координат не обязаны быть взаимно-перпендикулярными или близкими к этому.</p>	

### 2.3. Подготовка к интерактивному занятию №2 «Классификация на базе конструирования нейросети в специализированном нейропакете» (8 ч)

**Цель занятия:** активное воспроизведение ранее полученных знаний по разделу 2 «Методы классификации и кластеризации» в «незнакомых» условиях (применение основных понятий темы раздела 2 для решения задачи: апробация алгоритмов для практических задач).

**Дополнительная литература** для подготовки к занятию:

- 1) <http://www.alleng.ru/d/econ/econ292.htm> Юдин С.В. Математика в экономике. Тула: РГТЭУ, 2009. — 228 с.
- 2) [http://www.aup.ru/books/m155/2\\_12.htm](http://www.aup.ru/books/m155/2_12.htm) Орлов А.И. Математика случая. Вероятность и статистика – основные факты. Учебное пособие. М.: МЗ-Пресс, 2004.
- 3) Самостоятельный интернет-поиск.

**Форма текущего контроля** освоения компетенций (ОК-1, ОК-3, ОК-12, уровни З-Эл, У-Эл, В-Эл; ПК-26 уровни З-Пр, У-Пр, В-Пр) (см. табл.2 методических указаний к практическим занятиям): *отчет* по решению практических текстовых задач, *типовая формулировка* которых следующая:

**Задача И2.1.** Рассмотреть решение задачи «Выдавать ли кредит клиенту» в аналитическом специализированном нейропакете (например, Deductor, BaseGroup). В качестве обучающего набора данных выступает база данных, содержащая информацию о

Составитель: Колесникова С.И., каф.ЭМИС

клиентах, в частности: Сумма кредита, Срок кредита, Цель кредитования, Возраст, Пол, Образование, Частная собственность, Квартира, Площадь квартиры. На основе этих данных необходимо построить модель, которая сможет дать ответ, входит ли клиент, желающий получить кредит, в группу риска невозврата кредита, т.е. пользователь должен получить ответ на вопрос «Выдавать ли кредит клиенту?». Задача относится к группе задач классификации, т.е. обучения с учителем.

**Ознакомиться со следующим материалом (по указанным источникам):**

- 1) Персептрон Розенблатта.
- 2) Классификационные правила и их реализация в нейросети;
- 3) Возможности нейро-алгоритмов.
- 4) Примеры успешного решения нелинейных проблем на базе нейро-алгоритмов.

**Подготовить отчет команды, сформированной на ИнЗ 2, по обсуждаемым задачам, содержащий положения:**

- 1) Постановка решаемых задач.
- 2) Изложение обзора вариантов их решения.
- 3) Защита выбранного варианта и решение задач.
- 4) Подготовка презентации-защиты работы команды.

**Ознакомиться с формами текущего контроля**

Таблица 2

№ п / п	Наименование разделов	Формы контроля				
		Знаний	Умений	Навыков	Оценка личностных качеств	Компетенции/ ожидаемый уровень освоения
1	Случайные величины. Распределение вероятностей	Сдача индивидуальных заданий		Отчет по решению реальных практически задач на <i>интерактивном</i> занятии	Соблюдение установленных сроков для отчета и теста	ОК-1, ОК-3, ОК-12/ 3-Эл, У-Эл, В-Эл ПК-26/ 3-Пр, У-Пр, В-Пр
		Контрольная работа				
		Отчет по ИнЗ №2				

### 3. УКАЗАНИЯ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ по разделу 3

#### 3.1. Выполнение индивидуальных домашних заданий (ИДЗ) №3 по теме: «Методы построения математических моделей и прогнозирования временных рядов. Анализ свойств одномерных хаотических моделей»

*Цель занятия:* Проведение исследований на базе поискового метода по одной из заданных тем.

*Форма текущего контроля* освоения компетенций (ОК-1, ОК-3, ОК-12, уровни 3-Эл, У-Эл, В-Эл; ПК-26 уровни 3-Пр, У-Пр, В-Пр): отчет по выбранной теме и подготовка презентации-защиты:

*Рекомендуемые темы:*

- Метод модовой декомпозиции EMD (Empirical Mode Decomposition).
- Преобразование Гильберта-Хуанга ННТ (Hilbert-Huang Transform).
- Методы анализа стационарных стохастических временных рядов.

Составитель: Колесникова С.И., каф.ЭМИС

- Методы анализа нестационарных стохастических временных рядов.
- Типы нестационарности и методы обнаружения характера нестационарности.
- Нелинейная фильтрация сигнала с аддитивным шумом.
- Виды ядерных функций. Непараметрические оценки плотности вероятности типа Розенблатта – Парзена.
- Динамический хаос. Одномерные хаотические модели.

#### Дополнительная литература

1. Давыдов В.А., Давыдов А.В. Очистка геофизических данных от шумов с использованием преобразования Гильберта-Хуанга.// Электронное научное издание "Актуальные инновационные исследования: наука и практика", 2010, № 1. <http://www.actualresearch.ru>.
2. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование. – М.: Финансы и статистика, 2001. – 228 с.
3. Басвиль. М., Банвениста А. Обнаружение изменения свойств сигналов и динамических систем. Пер. с англ.- М.: Мир, 1989. - 278 с.
4. Безручко Б.П., Смирнов Д.А. Математическое моделирование и хаотические временные ряды. – Саратов: ГосУНЦ «Колледж», 2005. – 320 с.

### 3.2. Типовые тесты к разделу 3 (пример)

№	Вопросы	Отв еты
1.	<p><b>Регрессионная модель – это</b></p> <p>1) зависимость <math>f(x_k)=a \cdot y_k+b</math> на выборке <math>\{x_k, y_k\}_{k=1}^r</math>, для всех <math>k=1, \dots, r</math>;</p> <p>2) зависимость <math>A(x_k)=y_k</math> на выборке <math>\{x_k, y_k\}_{k=1}^r</math>, для всех <math>k=1, \dots, r</math>;</p> <p>3) функция установления степени соответствия набора <math>\{x_k, y_k\}_{k=1}^r</math> какой-либо функции из заданного набора..</p>	
2.	<p><b>Трендовый анализ — это:</b></p> <p>сравнение каждой позиции отчетности с предыдущим периодом определение структуры итоговых финансовых показателей с выявлением влияния каждой позиции отчетности на результат в целом сравнение каждой позиции отчетности с рядом предшествующих периодов расчет отношений между отдельными позициями отчета</p>	
3.	<p><b>Вертикальный (структурный) анализ – это:</b></p> <p>сравнение каждой позиции отчетности с предыдущим периодом сравнение каждой позиции отчетности с рядом предшествующих периодов расчет отношений между отдельными позициями отчета не подходит ни одно из приведенных здесь определений</p>	
4.	<p><b>Параметры в модели ARIMA(p, d, q) означают:</b></p> <p>1) большие нуля или равные нулю, и обозначают параметр авторегрессии, порядок интегрирования и скользящего среднего соответственно;</p> <p>2) произвольные целые числа, обозначающие параметр авторегрессии, порядок интегрирования и скользящего среднего соответственно;</p> <p>3) большие нуля или равные нулю, и обозначают точность модели,</p>	

	величину запаздывания и число слагаемых в модели.	
5.	<b>Расчет отношений между отдельными позициями отчета, определение взаимосвязей показателей – это:</b> горизонтальный (временной) анализ трендовый анализ вертикальный (структурный) анализ анализ коэффициентов	
6.	Ядерная функция – это....	
7.	Дисперсией случайного процесса $X(t)$ называется неслучайная функция $Dx(t)$ , которая при любом значении $t$ равна 1) математическому ожиданию соответствующего сечения случайного процесса 2) дисперсии соответствующего сечения случайного процесса 3) среднему квадратическому отклонению соответствующего сечения случайного процесса 4) вариации соответствующего сечения случайного процесса.	
8.	<b>Корреляционной функцией случайного процесса <math>X(t)</math></b> называется неслучайная функция $Kx(t, t')$ двух аргументов $t$ и $t'$ , которая при каждой паре значений $t$ и $t'$ равна 1) сумме математических ожиданий соответствующих сечений случайного процесса 2) сумме дисперсий соответствующих сечений случайного процесса 3) ковариации соответствующих сечений случайного процесса 4) произведению дисперсий соответствующих сечений случайного процесса	

### 3.3. Подготовка к интерактивному занятию №3 «Методы анализа нестационарных стохастических временных рядов»

**Цель занятия:** активное воспроизведение ранее полученных знаний по разделу 3 «Методы построения математических моделей и прогнозирования временных рядов» в «незнакомых» условиях (применение основных понятий темы раздела 3 для решения задачи: построение моделей прогнозирования для стохастических временных рядов).

**Форма текущего контроля** освоения компетенций (ОК-1, ОК-3, ОК-12, уровни З-Эл, У-Эл, В-Эл; ПК-26 уровни З-Пр, У-Пр, В-Пр) (см. табл.3 методических указаний к практическим занятиям): *отчет* по решению практических текстовых задач, *типовая формулировка* которых следующая:

**Задача ИЗ.1.** Типичной в сфере маркетинга является задача прогнозирования рынков. В результате решения данной задачи оцениваются перспективы развития конъюнктуры определенного рынка, изменения рыночных условий на будущие периоды, определяются тенденции рынка (структурные изменения, потребности покупателей, изменения цен). Обычно в этой области решаются следующие практические задачи:

- прогноз продаж товаров (например, с целью определения нормы товарного запаса);
- прогнозирование продаж товаров, оказывающих влияние друг на друга;
- прогноза продаж в зависимости от внешних факторов.

Для заданной *базы данных* (student\Колесникова\ Анализ данных) торговых временных рядов провести исследование по всем вышеперечисленным задачам на базе методов «Гусеница» и модовой декомпозиции. Сравнить результаты.

**Ознакомиться со следующим материалом (по указанным источникам):**

- 1) Методы главных компонент и «Гусеница»;
- 2) Метод модовой декомпозиции и преобразование Гильберта-Хуанга;

**Подготовить отчет команды, сформированной на ИнЗ 3, по обсуждаемым задачам, содержащий положения:**

- 1) Постановка решаемых задач.
- 2) Изложение обзора вариантов их решения.
- 3) Защита выбранного варианта и решение задач.
- 4) Подготовка презентации-защиты работы команды.

**Ознакомиться с формами текущего контроля**

Таблица 3

№ п / п	Наименование разделов	Формы контроля				
		Знаний	Умений	Навыков	Оценка личностных качеств	Компетенции/ ожидаемый уровень освоения
1	Основы теории случайных процессов	Сдача индивидуальных заданий		Отчет по решению реальных практических задач на интерактивном занятии	Соблюдение установленных сроков для отчета и теста	ОК-1, ОК-3, ОК-12/ 3-Эл, У-Эл, В-Эл ПК-26/ 3-Пр, У-Пр, В-Пр
		Контрольная работа				
		Отчет по ИнЗ №3				

#### 4. УКАЗАНИЯ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ по разделу 4

##### 4.1. Выполнение индивидуальных домашних заданий (ИДЗ) №4 по теме: «Основные модели управления данными, многомерный анализ данных»

*Цель занятия:* Проведение исследований на базе поискового метода по одной из заданных тем. Знакомство с основными моделями управления данными, с элементами многомерного анализа данных.

*Форма текущего контроля освоения компетенций компетенций (ОК-1, ОК-3, ОК-12, уровни 3-Эл, У-Эл, В-Эл; ПК-26 уровни 3-Пр, У-Пр, В-Пр):* отчет по выбранной теме и подготовка презентации-защиты:

*Рекомендуемые темы:*

- о СППР и их структура. Примеры СППР.
- о Комплексный подход к внедрению Data Mining, OLAP и хранилищ данных в СППР.
- о Основные методы многомерного анализа данных.
- о Гауссовский процесс и практические задачи.
- о Методы статистические в многомерном анализе данных.
- о Машинное обучение.
- о Задачи хаотической динамики и возможность управления хаосом.
- о Синергетический подход к управлению сложными объектами.

##### 4.2. Типовые тесты к разделу 4 (пример)

№	Вопросы	Отв
---	---------	-----

		еты
1.	По выборочным данным (5% отбор), удельный вес неуспевающих студентов на 4 курсе составил 5 %, а на 1 курсе – 15%. На каком курсе при одинаковой численности выборочной совокупности ошибка выборки больше?	
2.	Агент - это 1. Автономный искусственный объект, обладающий активным мотивированным поведением и способный к взаимодействию с другими объектами в динамических виртуальных средах. 2. Неавтономный искусственный объект, управляемый извне. 3. Неавтономный реальный объект, управляемый извне.	1
3.	Агенты, обладающие хорошо развитой и пополняемой символьной моделью внешнего мира благодаря наличию у них базы знаний, механизмов рассуждения и анализа действий, являются _(введите ответ)_ агентами	интеллектуальными
4.	Различают три типа многомерных OLAP-систем: 1. многомерный (Multidimensional) OLAP- MOLAP 2. реляционный (Relation) OLAP - ROLAP 3. смешанный или гибридный (Hibrid) OLAP – HOLAP. Более дешёвая реализация OLAP-системы по типу...	
5.	Знания это: 1) данные, информация, факты. 2) закономерности предметной области (принципы, связи, законы), полученные в результате практической деятельности и профессионального опыта, позволяющие специалистам ставить и решать задачи в этой области.	
6.	В каком виде чаще используются данные для анализа? 1. детальном 2. агрегированном	
7.	В двух третьих классах проводилось тестирование умственного развития десяти учащихся. Полученные значения величин средних достоверно не различались, однако психолога интересует вопрос — есть ли различия в степени однородности показателей умственного развития между классами. Варианты действий: 1. воспользоваться t-критерием Стьюдента; 2. по критерию Фишера необходимо сравнить дисперсии тестовых оценок в обоих классах; 3. сравнить на глазок (по процентным соотношениям) результаты;	
8.	Интеллектуальные базы данных характеризуются: 1. умением решать сложные плохо формализуемые задачи; 2. способностью к самообучению; 3. адаптивностью 4. неспособностью к самообучению (требуется «учитель»); 5. отсутствием адаптивности 6. способностью обеспечивать выборку необходимой информации, не	1-3,6

	присутствующей в явном виде, а выводимой из совокупности хранимых данных	
9.	<p>Для многоагентных систем характерны следующие особенности:</p> <ol style="list-style-type: none"> <li>1. Проведение альтернативных рассуждений на основе использования различных источников знаний с механизмом устранения противоречий;</li> <li>2. Распределенное решение проблем, которые разбиваются на параллельно решаемые подпроблемы, соответствующие самостоятельным источникам знаний;</li> <li>3. Применение множества стратегий работы механизма вывода заключений в зависимости от типа решаемой проблемы;</li> <li>4. Обработка больших массивов данных, содержащихся в базе данных;</li> <li>5. Использование различных математических моделей и внешних процедур, хранимых в базе моделей;</li> <li>6. Способность прерывания решения задач в связи с необходимостью получения дополнительных данных и знаний от пользователей, моделей, параллельно решаемых подпроблем.</li> </ol>	1-6

#### 4.3. Подготовка к интерактивному занятию №4 «Основные модели управления данными и распределённый анализ данных»

**Цель занятия:** активное воспроизведение ранее полученных знаний по разделу 4 «Основные модели управления данными, многомерный анализ данных» в «незнакомых» условиях (применение основных понятий темы раздела 4 «Основные модели управления данными, многомерный анализ данных» для решения задач: построение моделей анализа данных для практически важных текстовых задач).

**Форма текущего контроля** освоения компетенций (ОК-1, ОК-3, ОК-12, уровни З-Эл, У-Эл, В-Эл; ПК-26 уровни З-Пр, У-Пр, В-Пр) (см. табл.4 методических указаний к практическим занятиям): *отчет* по решению практических текстовых задач, *типовая формулировка* которых следующая:

**Задача И4.1.** Приведите примеры задача Data Mining в промышленном производстве:

- комплексный системный анализ производственных ситуаций;
- краткосрочный и долгосрочный прогноз развития производственных ситуаций;
- выработка вариантов оптимизационных решений;
- прогнозирование качества изделия в зависимости от некоторых параметров технологического процесса;
- обнаружение скрытых тенденций и закономерностей развития производственных процессов;
- прогнозирование закономерностей развития производственных процессов;
- обнаружение скрытых факторов влияния;
- обнаружение и идентификация ранее неизвестных взаимосвязей между производственными параметрами и факторами влияния;
- анализ среды взаимодействия производственных процессов и прогнозирование изменения ее характеристик;
- выработка оптимизационных рекомендаций по управлению производственными процессами;
- визуализация результатов анализа, подготовку предварительных отчетов и проектов допустимых решений с оценками достоверности и эффективности возможных реализаций.

Составитель: Колесникова С.И., каф.ЭМИС

**Задача И4.2.** Системы анализа распределённых данных. Разработка Web-сайта на основе базы данных.

*Темы для предварительного изучения.* Настройка Internet Information Server, знакомство с HTML, ASP, настройка DSN по эл. источнику: Сороковиков В. Н. Разработка Web-сайта на основе базы данных. – Томск. – 2002. (student\Колесникова\ Анализ данных)

*Опубликовать спроектированную БД согласно указанному алгоритму.*

**Ознакомиться со следующим материалом (по указанным источникам):**

- 1) Обзор систем анализа распределённых данных и принципы их построения;
- 2) Существующие стандарты Data mining и OLAP;
- 3) Понятие «мобильных агентов» и системы мобильных агентов, используемых для анализа данных;
- 4) Метод инвариантных многообразий для управления сложными системами.

**Подготовить отчет команды, сформированной на ИнЗ 4, по обсуждаемым задачам, содержащий положения:**

- 1) Постановка решаемых задач.
- 2) Изложение обзора вариантов их решения.
- 3) Защита выбранного варианта и решение задач.
- 4) Подготовка презентации-защиты работы команды.

**Ознакомиться с формами текущего контроля**

Таблица 4

№ п / п	Наименование разделов	Формы контроля				
		Знаний	Умений	Навыков	Оценка личностных качеств	Компетенции/ ожидаемый уровень освоения
1	Основные понятия математической статистики	Сдача индивидуальных заданий		Отчет по решению реальных практических задач на интерактивном занятии	Соблюдение установленных сроков для отчета и теста	ОК-1, ОК-3, ОК-12/ З-Эл, У-Эл, В-Эл ПК-26/ З-Пр, У-Пр, В-Пр
		Контрольная работа				
		Отчет по ИнЗ №3				

При составлении методических указаний использовался материал нижеуказанной литературы, а также материал интернет-ресурсов.

### ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. Васильев В.А. Непараметрическое оценивание функционалов от распределений стационарных последовательностей / В.А. Васильев, А.В. Добровидов, Г.М. Кошкин. – М.: Наука, 2004. – 508 с
2. Воронцов К.В. Лекции по методам оценивания и выбора моделей. 2007. Режим доступа: [www.ccas.ru/voron/download/Modeling.pdf](http://www.ccas.ru/voron/download/Modeling.pdf).
3. Воронцов К.В. Обзор современных исследований по проблеме качества обучения алгоритмов. Таврический вестник информатики и математики. – 2004. – № 1. – С. 5 – 24. <http://www.ccas.ru/frc/papers/voron04twim.pdf>.
4. Горелик А. Л., Скрипкин В. А. Методы распознавания: Учебное пособие для вузов. - 4-е изд., испр. - М.: Высшая школа, 2004. – 260 с.
5. Дюк В. Обработка данных на ПК в примерах / Вячеслав Дюк. - СПб.: Питер, 1997. - 240с.
6. З. Брандт. Анализ данных. Статистические и вычислительные методы для научных работников и инженеров: Пер. с англ.: Учебное пособие/ З. Брандт; пер.: О.И.Волкова; ред. пер.: Е. В. Чепурин. - М.: Мир, 2003 ; М.: АСТ, 2003. – 686 с.
7. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. - Новосибирск: ИМ СО РАН, 1999, стр. 270 [http://www.sernam.ru/book\\_zg.php](http://www.sernam.ru/book_zg.php)
8. Кокс, Д. Р. Анализ данных типа времени жизни : Пер. с англ. / Д. Р. Кокс, Д. Оукс; Пер. О. В. Селезнева, Ред. Ю. К. Беляев, Предисл. Ю. К. Беляев. - М. : Финансы и статистика, 1988. – 189с.
9. Лапко А.В. Непараметрические системы обработки информации : Учебное пособие для вузов / А. В. Лапко, С. В. Ченцов; Российская Академия наук. Сибирское отделение, Институт вычислительного моделирования. - М. : Наука, 2000. - 349 с.
10. Прикладная статистика. Классификация и снижение размерности : справочное издание / С. А. Айвазян [и др.] ; ред. С. А. Айвазян. - М. : Финансы и статистика, 1989. - 608 с.
11. Тюрин, Ю.Н.. Анализ данных на компьютере : учебное пособие для вузов / Ю. Н. Тюрин, А. А. Макаров. - 4-е изд., перераб. - М. : Форум, 2008. – 366.
12. <http://itteach.ru/predstavlenie-znaniy/metodi-i-sredstva-intellektualnogo-analiza-dannich><http://www.bsu.ru/content/hec/biometria/modules/stdatmin.html>
13. <http://www.arshinov74.ru/files/files/10.pdf>
14. <http://www.intuit.ru/department/database/dataanalysis/>
15. <http://www.olap.ru/basic/dm2.asp>
16. [http://www.sernam.ru/book\\_zg.php](http://www.sernam.ru/book_zg.php)
17. <http://www.statlab.kubsu.ru/node/4>