

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное учреждение  
высшего профессионального образования  
"Казанский (Приволжский) федеральный университет"  
Институт вычислительной математики и информационных технологий



**УТВЕРЖДАЮ**

Проректор  
по образовательной деятельности КФУ  
Проф. Таюрский Д.А.

"\_\_" \_\_\_\_\_ 20\_\_ г.

**Программа дисциплины**

Специальный семинар по анализу данных Б3.Б.16

Направление подготовки: 080500.62 - Бизнес-информатика

Профиль подготовки:

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

**Автор(ы):**

Миссаров М.Д.

**Рецензент(ы):**

Кашина О.А.

**СОГЛАСОВАНО:**

Заведующий(ая) кафедрой: Миссаров М. Д.

Протокол заседания кафедры No \_\_\_\_ от " \_\_\_\_ " \_\_\_\_\_ 201\_\_ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No \_\_\_\_ от " \_\_\_\_ " \_\_\_\_\_ 201\_\_ г

Регистрационный No

Казань  
2016

## Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) заведующий кафедрой, д.н. (доцент) Миссаров М.Д. кафедра анализа данных и исследования операций отделение фундаментальной информатики и информационных технологий , Moukadas.Missarov@kpfu.ru

### 1. Цели освоения дисциплины

Семинар охватывает следующие разделы:

- Корреляционный анализ;
- Регрессионный анализ;
- Дисперсионный анализ;
- Логические методы классификации;

### 2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б3.Б.16 Профессиональный" основной образовательной программы 080500.62 Бизнес-информатика и относится к базовой (общепрофессиональной) части. Осваивается на 4 курсе, 7 семестр.

Данный курс проводится на 4 курсе, в 7 семестре.

### 3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-12 (общекультурные компетенции)	осознает сущность и значение информации в развитии современного общества; владеет основными методами, способами и средствами получения, хранения, переработки информации;
ПК-20 (профессиональные компетенции)	использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования;
ПК-21 (профессиональные компетенции)	готовить научно-технические отчеты, презентации, научные публикации по результатам выполненных исследований.

В результате освоения дисциплины студент:

1. должен знать:

о современных методах Data Mining.

2. должен уметь:

понимать основные проблемы, возникающие при анализе данных, и пути их решения.

3. должен владеть:

навыками анализа данных различной природы.

-

### 4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) 72 часа(ов).

Форма промежуточного контроля дисциплины зачет в 7 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

#### 4.1 Структура и содержание аудиторной работы по дисциплине/ модулю Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Корреляционный анализ	7	1-3	0	6	0	устный опрос
2.	Тема 2. Регрессионный анализ.	7	4-8	0	10	0	устный опрос
3.	Тема 3. Дисперсионный анализ	7	9-11	0	6	0	устный опрос
4.	Тема 4. Логические методы классификации	7	12-16	0	10	0	коллоквиум
	Тема . Итоговая форма контроля	7		0	0	0	зачет
	Итого			0	32	0	

#### 4.2 Содержание дисциплины

##### Тема 1. Корреляционный анализ

###### *практическое занятие (6 часа(ов)):*

Основы корреляционного анализа Проверка значимости параметров связи Статистика Фишера

##### Тема 2. Регрессионный анализ.

###### *практическое занятие (10 часа(ов)):*

Однофакторная линейная регрессия. Множественная линейная регрессия. Проверка гипотез о коэффициентах регрессии. Нелинейная регрессия. Логистическая регрессия

##### Тема 3. Дисперсионный анализ

###### *практическое занятие (6 часа(ов)):*

Однофакторный дисперсионный анализ Двухфакторный дисперсионный анализ

##### Тема 4. Логические методы классификации

###### *практическое занятие (10 часа(ов)):*

Концептуальное обучение. Пути в пространстве гипотез. Деревья решений. Различные меры информативности. Обучение деревьев. Модели на основе правил.

#### 4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Корреляционный анализ	7	1-3	подготовка к устному опросу	8	устный опрос
2.	Тема 2. Регрессионный анализ.	7	4-8	подготовка к устному опросу	8	устный опрос
3.	Тема 3. Дисперсионный анализ	7	9-11	подготовка к устному опросу	12	устный опрос
4.	Тема 4. Логические методы классификации	7	12-16	подготовка к коллоквиуму	12	коллоквиум
	Итого				40	

#### 5. Образовательные технологии, включая интерактивные формы обучения

Обучение происходит в форме практических занятий и самостоятельной работы студентов. Список литературы разделен на две категории: необходимый для сдачи зачета минимум и дополнительная литература.

Изучение курса подразумевает не только овладение теоретическим материалом, но и получение практических навыков для более глубокого понимания разделов на основе решения задач и упражнений, иллюстрирующих доказываемые теоретические положения, а также развитие абстрактного мышления и способности самостоятельно доказывать утверждения. Самостоятельная работа предполагает выполнение домашних работ. Практические задания, выполненные в аудитории, предназначены для указания общих методов решения задач определенного типа. Закрепить навыки можно лишь в результате самостоятельной работы. Кроме того, самостоятельная работа включает подготовку к зачету. При подготовке к сдаче зачета весь объем работы рекомендуется распределять равномерно по дням, отведенным для подготовки к зачету, контролировать каждый день выполнения работы. Лучше, если можно перевыполнить план. Тогда будет резерв времени.

#### 6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

##### Тема 1. Корреляционный анализ

устный опрос , примерные вопросы:

Подготовка к семинарским занятиям: изучение теоретического материала, основной и дополнительной литературы по теме: -Представление и преобразование данных

##### Тема 2. Регрессионный анализ.

устный опрос , примерные вопросы:

Подготовка к семинарским занятиям: изучение теоретического материала, основной и дополнительной литературы по теме: -Кластерный анализ

##### Тема 3. Дисперсионный анализ

устный опрос , примерные вопросы:

Подготовка к семинарским занятиям: изучение теоретического материала, основной и дополнительной литературы по теме: -Байесовская классификация

#### **Тема 4. Логические методы классификации**

коллоквиум , примерные вопросы:

Подготовка к коллоквиуму по всем пройденным темам

#### **Тема . Итоговая форма контроля**

Примерные вопросы к зачету:

По данному курсу предусмотрен зачет.

Вопросы к зачету

1. Типы данных, обработка данных.
2. Меры сходства, метрики, ультраметрики.
3. Классификация задач анализа данных.
4. Эвристические алгоритмы кластерного анализа.
5. Иерархическая кластеризация .
6. Числовые характеристики кластерного разбиения.
7. Байесовский подход к задаче классификации.
8. Многомерное гауссовское распределение.
9. Линейный дискриминантный анализ.
10. Квадратичный дискриминантный анализ.
11. Наивный байесовский классификатор.
12. Однофакторная линейная регрессионный.
13. Множественная линейная регрессия.
14. Проверка гипотез о коэффициентах регрессии.
15. Нелинейная регрессия.
16. Логистическая регрессия

#### **7.1. Основная литература:**

1. Степанов, Роман Григорьевич. Технология Data Mining: Интеллектуальный анализ данных: учебное пособие / Р. Г. Степанов; Казан. гос. ун-т. Казань: Казанский государственный университет, 2009. 110 с
2. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. 3-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с.: ил. + CD-ROM (Учебная литература для вузов).  
<http://www.znanium.com/bookread.php?book=350638>
3. Компьютерные технологии анализа данных в эконометрике / Д.М. Дайитбегов. - 2-е изд., испр. и доп. - М.: Вузовский учебник: ИНФРА-М, 2010. - 578 с.: 70x100 1/16. - (Научная книга). (переплет) ISBN 978-5-9558-0191-9 <http://www.znanium.com/bookread.php?book=251791>
4. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. - (Университетская серия). - ISBN 978-5-4257-0092-6.  
<http://www.znanium.com/bookread.php?book=451186>
5. Кашина О.А., Миссаров М.Д. Электронный курс "Анализ данных в среде R", 2013  
<http://zilant.kpfu.ru/course/view.php?id=17341>
6. Кашина О.А., Миссаров М.Д. Электронный курс "Статистический анализ данных", 2013  
<http://zilant.kpfu.ru/course/view.php?id=17260>

## 7.2. Дополнительная литература:

1. Интеллектуальный анализ временных рядов[Электронный ресурс]: Учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - М.: ИД ФОРУМ: ИНФРА-М, 2012. - 160 с. . - Режим доступа: <http://www.znaniium.com/bookread.php?book=2493141>
2. Анализ данных : стат. и вычисл. методы для науч. работников и инженеров / З. Брандт ; Пер. с англ. О.И. Волковой ; Под ред. Е.В. Чепурина .? Москва : МИР : АСТ, 2003 .? 686 с. : ил. ; 24 + 1 опт. диск (CD-ROM) .? Загл. и авт. ориг.: Data Analysis / Siegmund Brandt .? Библиогр. в подстроч. примеч. ? Указ. компьютер. прогр.: с. 666-667 .? Предм. указ.: с. 668-674 .? ISBN 5-03-003478-1 (Мир) .? ISBN 5-17-019778-0 (АСТ) , 5000.
3. Геоэкономический анализ данных в экологии и природопользовании (с применением пакета R) : учебное пособие / Казан. (Приволж.) федер. ун-т, Ин-т экологии и географии ; [авт.-сост.: д-р биол. наук, проф. А. А. Савельев и др.] .? Казань : Казанский университет, 2012 .? 120 с. : ил. ; 21 .? Библиогр.: с. 108-110 (22 назв.).
4. Анализ данных секвенирования транскриптома и метаболома [Текст: электронный ресурс] : учебно-методическое пособие / Н. И. Акберова .? Электронные данные (1 файл: 1,19 Мб) .? (Казань : Казанский федеральный университет, 2014) .? Загл. с экрана .? Для 3-го семестра .? Вых. дан. ориг. печ. изд.:.  
Режим доступа: только для студентов и сотрудников КФУ .?  
<URL:[http://libweb.kpfu.ru/ebooks/01-IFMB/01\\_012\\_000740.pdf](http://libweb.kpfu.ru/ebooks/01-IFMB/01_012_000740.pdf)>

## 7.3. Интернет-ресурсы:

- Data Mining Labs - <http://dmlabs.org/>  
The R Project for Statistical Computing - <http://www.r-project.org/>  
курс "Data mining" - <http://www.intuit.ru/studies/courses/6/6/info>  
курс "Интеллектуальный анализ данных средствами MS SQL Server 2008" - <http://www.intuit.ru/studies/courses/2312/612/info>  
программный комплекс RStudio для интеллектуального анализа данных - <http://www.rstudio.com/>

## 8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Специальный семинар по анализу данных" предполагает использование следующего материально-технического обеспечения:

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, УМК, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

Практические занятия по дисциплине проводятся в аудитории, оснащенной доской и мелом (маркером).

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 080500.62 "Бизнес-информатика" .

Автор(ы):

Миссаров М.Д. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.

Рецензент(ы):

Кашина О.А. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.