

Документ подписан простой электронной подписью

Информация о документе:

ФИО: Хоружий Л.И. Хоружий

Должность: Директор института экономики и управления АПК

Дата подписания: 13.07.2023 19:20:34

Уникальный программный ключ:

1e90b132d9b04dce67585160b015dddf2cb1e6a9



**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ**

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ –**

**МСХА имени К.А. ТИМИРЯЗЕВА»**

**(ФГБОУ ВО РГАУ - МСХА имени К.А. Тимирязева)**

Институт экономики и управления АПК  
Кафедра статистики и кибернетики

УТВЕРЖДАЮ:  
Директор института экономики и управ-  
ления АПК  
Л.И. Хоружий  
“ 13 ” 07 2022 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**  
**Б1.В.ДВ.06.01 Интеллектуальный анализ данных и статистика на ино-**  
**странном языке**

для подготовки бакалавров

ФГОС ВО

Направление: 09.03.02 «Информационные системы и технологии»

Направленность:

Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)

Курс 4

Семестр 7,8

Форма обучения очная

Год начала подготовки 2022

Москва, 2022

Разработчик (и): Харитонов А.Е., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

  
«26» 08 2022г.

(подпись)

Рецензент: Коломеева Е.С., к.э.н.

(ФИО, ученая степень, ученое звание)

«26» 08 2022г.

Программа составлена в соответствии с требованиями ФГОС ВО, профессионального стандарта и учебного плана по направлению подготовки 09.03.02 «Информационные системы и технологии».

Программа обсуждена на заседании кафедры статистики и кибернетики протокол № 11 от «26» августа 2022 г.

И.о.зав. кафедрой Уколова А.В., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

  
(подпись)

«26» 08 2022г.

**Согласовано:**

Председатель учебно-методической комиссии института экономики и управления АПК  
Корольков А.Ф., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

  
(подпись)

«26» 08 2022г.

И.о.заведующего выпускающей кафедрой  
статистики и кибернетики

Уколова А.В., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

  
(подпись)

«26» 08 2022г.

/Заведующий отделом комплектования ЦНБ

  
(подпись)

## СОДЕРЖАНИЕ

<b>АННОТАЦИЯ .....</b>	<b>4</b>
<b>1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....</b>	<b>5</b>
<b>2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ .....</b>	<b>5</b>
<b>3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....</b>	<b>5</b>
<b>4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ .....</b>	<b>5</b>
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ .....	6
ПО СЕМЕСТРАМ .....	6
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ .....	11
4.3 ПРАКТИЧЕСКИЕ ЗАНЯТИЯ .....	15
<b>5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ .....</b>	<b>21</b>
<b>6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....</b>	<b>21</b>
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ .....	22
6.2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ.....	39
<b>7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....</b>	<b>40</b>
7.1 ОСНОВНАЯ ЛИТЕРАТУРА.....	40
7.2 ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА .....	<b>Ошибка! Закладка не определена.</b>
7.3 МЕТОДИЧЕСКИЕ УКАЗАНИЯ, РЕКОМЕНДАЦИИ И ДРУГИЕ МАТЕРИАЛЫ К ЗАНЯТИЯМ.....	40
<b>8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ .....</b>	<b>41</b>
<b>9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ.....</b>	<b>41</b>
<b>10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ .....</b>	<b>42</b>
<b>11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ..</b>	<b>43</b>
<b>12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ.....</b>	<b>43</b>

## Аннотация

### рабочей программы учебной дисциплины

#### **Б1.В.ДВ.06.01 «Интеллектуальный анализ данных и статистика на иностранном языке»**

**для подготовки бакалавров по направлению 09.03.02 «Информационные системы и технологии» по направленности «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»**

**Цель освоения дисциплины:** Целью дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в проведении анализ данных с использованием информационных технологий и навыками чтения и перевода текстов на иностранном языке в профессиональном общении.

**Место дисциплины в учебном плане:** включена в часть, формируемую участниками образовательных отношений (дисциплина по выбору) учебного плана по направлению подготовки 09.03.02 «Информационные системы и технологии».

**Требования к результатам освоения дисциплины:** в результате освоения дисциплины формируются следующие компетенции (индикаторы): УК-4(УК-4.2, УК-4.3); ПКос-8 (ПКос-8.1, ПКос-8.2, ПКос-8.3); ПКос-9 (ПКос-9.2, ПКос-9.3)

#### **Краткое содержание дисциплины:**

The Subject and the Method of Statistics. Absolute, Mean and Relative Values. Testing Statistical Hypotheses. Definition, Essence and Tasks of Data Mining. The Phases of Data Mining Project. Solving the Problems of Classification and Regression. Search for Association Rules during Data Analysis. Cluster Analysis. Text Mining. Data Visualization.

Предмет и метод статистики. Абсолютные, средние и относительные величины. Проверка статистических гипотез. Понятие, сущность и задачи интеллектуального анализа данных. Стадии интеллектуального анализа данных. Решение задач классификации и регрессии. Поиск ассоциативных правил в процессе анализа данных. Кластерный анализ. Анализ текстовой информации и аналитика. Визуализация данных.

**Общая трудоемкость дисциплины составляет:** 5 зачетных единиц (180 часов).

**Промежуточный контроль:** зачет в 7 семестре, экзамен в 8 семестре.

## **1. Цель освоения дисциплины**

Целью дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в проведении анализ данных с использованием информационных технологий и навыками чтения и перевода текстов на иностранном языке в профессиональном общении.

## **2. Место дисциплины в учебном процессе**

Дисциплина «Интеллектуальный анализ данных и статистика на иностранном языке» включена в часть, формируемую участниками образовательных отношений (дисциплина по выбору) учебного плана. Дисциплина «Интеллектуальный анализ данных и статистика на иностранном языке» реализуется в соответствии с требованиями ФГОС ВО, профессионального стандарта, ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии».

Предшествующими курсами, на которых непосредственно базируется дисциплина «Интеллектуальный анализ данных и статистика на иностранном языке» являются «Линейная алгебра», «Математический анализ», «Математическая статистика», «Теория вероятностей», «Алгоритмизация и программирование», «Моделирование временных рядов на иностранном языке», «Моделирование временных рядов», «Анализ экономических данных с использованием современных информационных технологий на иностранном языке», «Анализ экономических данных с использованием современных информационных технологий», «Английский язык для IT-специалистов», «Адаптационный курс английского языка», «Расчетно-графическая статистика в среде R на иностранном языке», «Расчетно-графическая статистика в среде R».

Рабочая программа дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.

## **3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы**

Образовательные результаты освоения дисциплины обучающимся, представлены в таблице 1.

## **4. Структура и содержание дисциплины**

#### **4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам**

Общая трудоёмкость дисциплины составляет 5 зач.ед. (180 часов), их распределение по видам работ семестрам представлено в таблице 2.

## Требования к результатам освоения учебной дисциплины

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
1.	УК-4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)	УК-4.2 Уметь: применять на практике деловую коммуникацию в устной и письменной формах, методы и навыки делового общения на русском и иностранном языках		вести монологическую и диалогическую речь в рамках изученных тем с учетом правил речевого общения в информационно-технологической и статистической сфере на русском и иностранном языках	
2.			Владеть: навыками чтения и перевода текстов на иностранном языке в профессиональном общении; навыками деловых коммуникаций в устной и письменной форме на русском и иностранном языках; методикой составления суждения в межличностном деловом общении на русском и иностранном языках			навыками чтения и перевода текстов на иностранном языке, относящимися к информационно-технологической и статистической деятельности
3.	ПКос-8	Способность проводить анализ данных с использованием информационных технологий в области сельского хозяйства, экономики, бухгалтерского учета, статисти-	ПКос-8.1 Знать: основы технологии производства продукции сельского хозяйства; теорию и методологию дисциплин экономического профиля (экономика, бухгалтерский учет, статистика, финансы и др.); информационные технологии анализа данных; источники информации для профес-	Современные программные средства интеллектуального анализа данных спо-		

		стики, финансов и др.	сиональной деятельности	собы и средства сбора статистической информации для профессиональной деятельности		
4.			ПКос-8.2 Уметь: собирать информацию для проведения анализа; устанавливать причинно-следственные связи между признаками; выбирать и применять, в том числе с использованием современных информационных технологий, методы анализа данных; делать выводы на основе проведенного анализа данных		применять информационные технологии и статистические методы для сбора, обработки и анализа данных, устанавливать причинно-следственные связи между признаками с использованием статистических методов и интеллектуального анализа данных; интерпретировать полученные результаты	
5.			ПКос-8.3 Владеть: методологией и навыками проведения анализа данных с использованием информационных технологий в области экономики, в том числе экономики сельского хозяйства			навыками разработки информационных систем обработки данных в области экономики, в том числе экономики сельского хозяйства
6.	ПКос-9	Способен осуществлять разработку, от-	ПКос-9.2 Уметь: использовать методы и приемы формализации и алгоритми-		применять языки программирования	



		ладку и рефакторинг кода программного обеспечения, интеграцию программных модулей и компонент, в том числе взаимодействующих с внешней средой, средствами выбранных языков программирования	зации поставленных задач; использовать программные продукты для графического отображения алгоритмов; применять стандартные алгоритмы в соответствующих областях; применять выбранные языки программирования для написания программного кода; использовать выбранную среду программирования; применять инструментарий для создания и актуализации исходных текстов программ; выявлять ошибки в программном коде, интерпретировать сообщения об ошибках, предупреждения, записи технологических журналов; применять методы и приемы отладки программного кода		R и Python для написания программного кода для проведения интеллектуального анализа данных	
7.			ПКос-9.3 Владеть навыками; составления формализованных описаний решений поставленных задач в соответствии с требованиями технического задания; разработки алгоритмов решения поставленных задач в соответствии с требованиями технического задания или других принятых в организации нормативных документов; создания программного кода в соответствии с техническим заданием (готовыми спецификациями); оптимизации программного кода с использованием специализированных программных средств; анализа и проверки исходного программного кода; отладки программного кода на уровне про-			навыками создания программного кода в соответствии с техническим заданием

			граммных модулей и межмодульных взаимодействий и взаимодействий с окружением			
--	--	--	--	--	--	--

## Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость всего/*		
	час.	в т.ч. по семестрам	
		№ 7	№ 8
<b>Общая трудоёмкость</b> дисциплины по учебному плану	<b>180,00/8</b>	<b>72,00/4</b>	<b>108,00/4</b>
<b>1. Контактная работа:</b>	<b>100,65/8</b>	<b>50,25/4</b>	<b>50,40/4</b>
<b>Аудиторная работа</b>	<b>100,65/8</b>	<b>50,25/4</b>	<b>50,40/4</b>
<i>в том числе:</i>			
<i>лекции (Л)</i>	40,00	16,00	24,00
<i>практические занятия (ПЗ)</i>	58,00/8	34,00/4	24,00/4
<i>консультации перед экзаменом</i>	2,00	-	2,00
<i>контактная работа на промежуточном контроле (КРА)</i>	0,65	0,25	0,40
<b>2. Самостоятельная работа (СРС)</b>	<b>79,35</b>	<b>21,75</b>	<b>57,60</b>
<i>контрольная работа</i>	15,00	5,00	10,00
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к практическим занятиям и т.д.)</i>	30,75	7,75	23,00
<i>Подготовка к зачету</i>	9,00	9,00	-
<i>Подготовка к экзамену</i>	24,60	-	24,60
Вид промежуточного контроля:		зачёт	экзамен

\* в том числе практическая подготовка.

## 4.2 Содержание дисциплины

## Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнёно)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	всего/*	ПКР	
<b>Section 1. Statistics</b> <b>Раздел 1. Статистика</b>	<b>25,75</b>	<b>4</b>	<b>16</b>		<b>5,75</b>
Theme 1. Statistics and data analysis Тема 1. Статистика и анализ данных	<b>3,75</b>	1	2		0,75
Theme 2. Absolute, Mean and Relative Values Тема 2. Абсолютные, средние и относительные величины	<b>4</b>	1	2		1
Theme 3. Testing Statistical Hypotheses Тема 3. Проверка статистических гипотез.	<b>12</b>	2	8		2
Theme 4. Correlation and regression analysis Тема 4. Корреляционно-регрессионный анализ	<b>6</b>	-	4		2
<b>Section 2. Data Mining</b> <b>Раздел 2. Интеллектуальный анализ данных</b>	<b>46</b>	<b>12</b>	<b>18</b>		<b>16</b>
Theme 5. Definition, Essence and Tasks of Data Mining Тема 5. Понятие, сущность и задачи интел-	<b>10</b>	2	2		6

лектуального анализа данных					
Theme 6. Classification and Regression Task Тема 6. Решение задач классификации и регрессии	36/4	10	16/4		10
Контактная работа на промежуточном контроле (КРА)	0,25			0,25	
<b>Всего за 7 семестр</b>	<b>72/4</b>	<b>16</b>	<b>34</b>	<b>0,25</b>	<b>21,75</b>
Theme 7. Association Rule Mining Тема 7. Поиск ассоциативных правил в процессе анализа данных	22	6	4		12
Theme 8. Cluster Analysis Тема 8. Кластерный анализ	18/4	6	6/4		6
Theme 9. Neural network analysis methods Тема 9. Нейросетевые методы анализа	26,6	4	6		16,6
Theme 10. Text Mining Тема 10. Анализ текстовой информации и аналитика	20	4	4		12
Theme 11. Data Visualization Тема 11. Визуализация данных	19	4	4		11
Консультации перед экзаменом	2			2	
Контактная работа на промежуточном контроле (КРА)	0,4			0,4	
<b>Всего за 8 семестр</b>	<b>108/4</b>	<b>24</b>	<b>24/4</b>	<b>2,4</b>	<b>57,6</b>
<b>Итого по дисциплине</b>	<b>180/8</b>	<b>40</b>	<b>58/8</b>	<b>2,65</b>	<b>79,35</b>

\* в том числе практическая подготовка

## Section 1. Statistics

### Раздел 1 Статистика

#### Theme 1. Statistics and data analysis

Statistics as a science. The subject of statistics. The method of statistics. Statistical regularity. The main concepts of statistics: statistical population, population unit, attribute, variance, statistical indicator, system of indicators. Differences between data mining and statistics.

#### Тема 1. Статистика и анализ данных

Понятие о статистике как науке. Предмет статистики. Метод статистики. Статистическая закономерность. Основные категории и понятия статистики: статистическая совокупность, единица совокупности, признак, вариация, статистический показатель, система показателей. Различия между интеллектуальным анализом данных и статистикой.

#### Theme 2. Absolute, Mean and Relative Values

Significance of absolute and relative values for statistical analysis. Types of absolute values, their measurement units and ways of obtaining. Mean value in statistics, its essence and conditions of applying. Relative values, their types and ways of calculating in packages of statistical programs. Relationship between absolute and relative values.

#### Тема 2. Абсолютные, средние и относительные величины

Значение абсолютных и относительных величин для статистического анализа данных. Виды абсолютных величин, единицы измерения и способы получения. Средняя величина в статистике, ее сущность и условия применения. Относительные величины, их виды и способы расчета в пакетах статистических программ. Взаимосвязь абсолютных и относительных величин.

### **Theme 3. Testing Statistical Hypotheses**

Definition of sample observation. Theoretical concepts of sample observation. General population and sampling, their unifying characteristics. Mean and maximum error of sampling. Definition of statistical hypotheses. The main criteria of testing statistical hypotheses. Phases of testing statistical hypotheses. Student's t-test and F-test. Wilcoxon rank correlation test. Test Shapiro-Wilks. Kolmogorov-Smirnov test. Variance analysis.

### **Тема 3. Проверка статистических гипотез.**

Понятие выборочного наблюдения. Теоретические основы выборочного наблюдения. Генеральная и выборочная совокупности, их обобщающие характеристики. Средняя и предельная ошибки выборки. Понятие статистической гипотезы. Основные критерии для проверки статистических гипотез. Этапы проверки статистических гипотез. Критерий t-Стьюдента и F-Фишера. Тест ранговой корреляции Уилкоксона. Тест Шапиро-Уилкса. Тест Колмогорова-Смирнова. Дисперсионный анализ.

### **Theme 4. Correlation and regression analysis**

The general equation of pair and multiple linear regression and their selective estimates. Interpretation of the parameters of the regression equation. Model quality indicators. Estimation of the reliability of parameters and correlation coefficient using Student's criterion. Estimation of the reliability of the regression equation in general, analysis of variance, Fisher criterion.

Prediction based on the regression equation. Construction of nonlinear regressions in R. environment. Estimation of the reliability of parameters of nonlinear models. Selection of factors in a multiple regression model. The version of all possible regressions and step-by-step selection procedures in R.

### **Тема 4. Корреляционно-регрессионный анализ**

Генеральное уравнение парной и множественной линейной регрессии и их выборочные оценки. Интерпретация параметров уравнения регрессии. Показатели качества модели. Оценка достоверности параметров и коэффициента корреляции с использованием критерия Стьюдента. Оценка достоверности уравнения регрессии в целом, дисперсионный анализ, критерий Фишера.

Прогнозирование на основе уравнения регрессии. Построение нелинейных регрессий в среде R. Оценка достоверности параметров нелинейных моделей. Отбор факторов в модель множественной регрессии. Версия всех возможных регрессий и пошаговые процедуры отбора в среде R.

## **Section 2. Data Mining**

## **Раздел 2. Интеллектуальный анализ данных**

### **Theme 5. Definition, Essence and Tasks of Data Mining**

History, basic concepts and main tasks of data mining. Perspective on the data mining technology as a part of the information technology market. Phases of data mining: defining the aims, collecting the data, database creation, exploring and preparing the data, population segmentation, drawing up and validating the predictive models, deploying the models, training the model users, monitoring and enriching the models.

### **Тема 5. Понятие, сущность и задачи интеллектуального анализа данных**

История возникновения, основные понятия и главные задачи интеллектуального анализа данных. Взгляд на технологию Data Mining как на часть рынка информационных технологий. Стадии Data Mining: постановка цели и задач, сбор данных, создание базы данных, исследование и подготовка данных, разделение совокупности на сегменты, построение и проверка достоверности предсказательной модели, инсталляция модели, обучение пользователей, использующих модель, мониторинг, расширение модели.

### **Theme 6. Classification and Regression Task**

Classification rules and mathematical functions and methods of their construction. Construction of the main classification methods in the environment of R. Decision tree. Method support. Random forest. Logistic regression. Discriminant analysis. Bayesian (naive) classification. Closest neighbor method.

### **Тема 6. Решение задач классификации и регрессии**

Правила классификации и математические функции и методы их построения. Построение основных методов классификации в среде R. Дерево решений. Метод опорных. Случайный лес. Логистическая регрессия. Дискриминантный анализ. Байесовская (наивная) классификация. Метод ближайшего соседа.

### **Theme 7. Association Rule Mining**

Concepts and methods of identifying patterns in data mining. Applying the discovered patterns for prediction of unknown values (predictive modeling). Analysis of exceptions designed to detect and interpret anomalies in the discovered patterns.

### **Тема 7. Поиск ассоциативных правил в процессе анализа данных**

Понятия и методы выявления закономерностей в интеллектуальном анализе данных. Использование выявленных закономерностей для предсказания неизвестных значений (прогностическое моделирование). Анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях.

### **Theme 8. Cluster Analysis**

Hierarchical and non-hierarchical algorithms of cluster analysis. Examples of the applying the cluster analysis.

### **Тема 8. Кластерный анализ**

Иерархические и неиерархические методы в кластерном анализе. Рассмотрение примеров использования кластерного анализа.

### **Theme 9. Neural network analysis methods**

The essence and application of neural network analysis methods. Classifiers based on artificial neural networks. Setting parameters of neural networks using the caret package.

#### **Тема 9. Нейросетевые методы анализа**

Сущность и применение нейросетевых методов анализа. Классификаторы на основе искусственных нейронных сетей. Настройка параметров нейронных сетей средствами пакета caret.

#### **Theme 10. Text Mining**

Basic methods of text mining and analysis of text data to detect patterns with an emphasis on statistical approaches.

#### **Тема 10. Анализ текстовой информации и аналитика**

Основные методы добычи и анализа текстовых данных с целью обнаружения закономерностей с акцентом на статистические подходы.

#### **Theme 11. Data Visualization**

Methods and means of visual presentation of information, in particular, ways of presenting information in one-, two-, three-dimensional spaces and ways of displaying information in more than three dimensions. The description of the principles of qualitative visualization. The main trends in data visualization.

#### **Тема 11. Визуализация данных**

Методы и средства визуального представления информации, в частности, способы представления информации в одно-, двух-, трехмерном измерениях, а также способы отображения информации в более чем трех измерениях. Описание принципов качественной визуализации. Основные тенденции в области визуализации.

### **4.3 Лекции/ практические занятия**

Таблица 4

#### **Содержание практических занятий и контрольные мероприятия**

<b>№ п/п</b>	<b>№ раздела</b>	<b>№ и название лекций/ практических занятий</b>	<b>Формируемые компетенции (индикаторы)</b>	<b>Вид контрольного мероприятия</b>	<b>Кол-во часов/ из них практическая подготовка</b>
1.	<b>Section 1. Statistics</b>				
	<b>Раздел 1. Статистика</b>				
	Theme 1. Statistics and data analysis Тема 1. Статистика и анализ данных	Лекция № 1. Программные средства статистической обработки данных Lecture number 1. Software statistical data processing	УК-4.2 УК -4.3 ПКос-8.1	-	1
		Практическая работа № 1 Предмет и метод и задачи статистики Practice Class 1 The subject and method and objectives of statistics	УК-4.2 УК -4.3 ПКос-8.1	Чтение, перевод, участие в обсуждении	2

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
	Theme 2. Absolute, Mean and Relative Values Тема 2. Абсолютные, средние и относительные величины	Лекция № 2. Возможности среды R для обработки данных Lecture number 2. The possibilities of the medium R for data processing	УК-4.2 УК -4.3 ПКос-8.1	-	1
		Практическое занятие №2 Расчет основных показателей центральной тенденции и вариации в пакете R» Practice Class 2 Calculation of the main indicators of the central trend and variation in the package R	УК-4.2 УК -4.3 ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Чтение, перевод, защита работы	2
	Theme 3. Testing Statistical Hypotheses Тема 3. Проверка статистических гипотез.	Лекция № 3. Статистические тесты, реализованные в пакетах R и Python Lecture number 3. Statistical tests implemented in packages R and Python	УК-4.2 УК -4.3 ПКос-8.1	-	2
		Практическое занятие № 3. Статистические тесты Practical lesson № 3. Statistical tests in the environment	УК-4.2 УК -4.3 ПКос-8.2 ПКос-8.3	Чтение, перевод, защита работы	4
		Практическое занятие № 4. Разработка статистических тестов на языке R Practical lesson № 4. Development of statistical tests in the R language	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	4
	Theme 4. Correlation and regression analysis Тема 4. Корреляционно-регрессионный анализ	Практическое занятие № 5. Корреляционно-регрессионный анализ Practical lesson number 5. Correlation and regression analysis in the environment	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Контрольная работа по темам 1, 2, 3 и 4. Examination on topics 1, 2, 3 and 4.	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Письменная работа	2
2	<b>Section 2. Data Mining</b> <b>Раздел 2. Интеллектуальный анализ данных</b>				
	Theme 5. Definition, Essence and Tasks of Data Mining Тема 5. Понятие,	Лекция № 4. Реализация моделей Data mining в среде R и Python Lecture No. 4. Implementation of Data mining models in the R en-	УК-4.2 УК -4.3 ПКос-8.1	-	2



№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
	сущность и задачи интеллектуального анализа данных (название)	vironment and Python			
		Практическое занятие №6 «Интеллектуальный анализ данных и статистика: в чём различия? Practice Class 6 «Data mining and statistics: how are they different»?	УК-4.2 УК -4.3	Чтение, перевод, участие в обсуждении	2
	Theme 6. Classification and Regression Task Тема 6. Решение задач классификации и регрессии	Лекция № 5. Классификация методов Data mining Lecture number 5. Classification of methods of data mining	УК-4.2 УК -4.3 ПКос-8.1		2
		Лекция № 6. Методы классификации в Data mining Lecture number 6. Classification methods in Data mining	УК-4.2 УК -4.3 ПКос-8.1		6
		Лекция № 7. Регрессия в Data mining Lecture number 7. Regression in Data mining	УК-4.2 УК -4.3 ПКос-8.1		2
		Практическое занятие №7 Построение деревьев решений Practical lesson №7 Construction of decision trees	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2/2
		Практическое занятие №8 Метод опорных векторов Practical lesson №8 Method of support vectors	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2/2
		Практическое занятие №9 Случайный лес Practical lesson №9 Random forest	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Практическое занятие №10 Логистическая регрессия Practical lesson №10 Logistic regression	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Практическое занятие №11 Дискриминантный анализ Practical lesson №11 Discriminant analysis	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Практическое занятие №12 Байесовская (наивная) классификация	ПКос-8.2 ПКос-8.3 ПКос-9.2	Защита работы	2

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
		Practical lesson №12 Bayesian (naive) classification	ПКос-9.3		
		Практическое занятие №13 Метод ближайшего соседа Practical lesson №13 Metod nearest neighbor	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Контрольная работа по темам 5 и 6 Examination of topics 5 and 6	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Контрольная работа (письменная работа)	2
	Theme 7. Association Rule Mining Тема 7. Поиск ассоциативных правил в процессе анализа данных	Лекция 8. Поиск ассоциативных правил в процессе анализа Lecture 8. Search for associative rules in the analysis process	УК-4.2 УК -4.3 ПКос-8.1	-	6
		Практическое занятие №14 «Методы поиска ассоциативных правил» Practice Class 14 «Methods of association rule mining»	УК-4.2 УК -4.3 ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Чтение, перевод, защита работы	2
		Практическое занятие №15 «Алгоритм Apriori» Practice Class 15 «Apriori algorithm»	УК-4.2 УК -4.3	Чтение, перевод	2
	Theme 8. Cluster Analysis Тема 8. Кластерный анализ	Лекция 8. Кластерный анализ в среде R Lecture 8. Cluster analysis R	УК-4.2 УК -4.3 ПКос-8.1	-	6
		Практическое занятие №16 «Иерархические методы в кластерном анализе» Practice class 16 «Hierarchical algorithms of cluster analysis»	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2/2
		Практическое занятие №17 «Неиерархические методы в кластерном анализе» Practice class 17 «Non-hierarchical algorithms of cluster analysis»	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	4/2
	Тема 9. Нейросетевые методы анализа	Лекция 9. Нейросетевые методы анализа Lecture 9. Neural Network Analysis Methods	УК-4.2 УК -4.3 ПКос-8.1	-	4
		Практическое занятие №18 Классификаторы на основе искусственных нейронных сетей Practical lesson №18 Classifiers	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	3

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
		based on artificial neural networks			
		Практическое занятие №19 Настройка параметров нейронных сетей средствами пакета Caret Practical lesson №19 Setting parameters of neural networks by means of the caret package	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	3
	Theme 10. Text Mining Тема 10. Анализ текстовой информации и аналитика	Лекция 8. Анализ текстовой информации и аналитика Lecture 8. Text information analysis and analytics	УК-4.2 УК -4.3 ПКос-8.1	-	4
		Практическое занятие №20 «Анализ текстовой информации: подходы, особенности, перспективы» Practice Class 20 «Text Mining: approaches, features, perspectives»	УК-4.2 УК -4.3	Чтение, перевод, участие в обсуждении	2
		Практическое занятие №21 «Типичные задачи анализа текстовой информации: категоризация, кластеризация текстов, смысловой анализ» Practice Class 21 «Typical tasks of text mining: categorization, text clustering, sentiment analysis»	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Лекция 11. Методы и средства визуализации данных Lecture 11. Methods and tools for data visualization	УК-4.2 УК -4.3 ПКос-8.1		4
	Theme 11. Data Visualization Тема 11. Визуализация данных	Практическое занятие №22 «Классификация методов визуализации данных» Practice Class 22 «Classification of methods of data visualization»	УК-4.2 УК -4.3 ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Участие в обсуждении Защита работы	1
		Практическое занятие №23 «Программное обеспечение для визуализации данных» Practice Class 23 «Data visualization software»	УК-4.2 УК -4.3	Чтение, перевод, участие в обсуждении	1
		Итоговая контрольная работа	УК-4.2	Тестирование	2

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
		The final test	УК -4.3 ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3		

Таблица 5

**Перечень вопросов для самостоятельного изучения дисциплины**

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
<b>Section 1. Statistics</b>		
Раздел 1. Статистика		
1.	Theme 1. Statistics and data analysis Тема 1. Статистика и анализ данных	Law of large numbers and its significance for statistics Statistical observation and its organization in foreign countries Закон больших чисел и его значение для статистики Статистическое наблюдение и его организация в зарубежных странах (УК -4.3, УК -4.3, ПКос-8.1)
2.	Theme 2. Absolute, Mean and Relative Values Тема 2. Абсолютные, средние и относительные величины	Conditions of application of relative values Arithmetic mean and its properties Mode, median, quartiles Условия применения относительных величин Среднее арифметическое и его свойства Мода, медиана, квартили (ПКос-8.1, ПКос-8.2)
3	Theme 3. Testing Statistical Hypotheses Тема 3. Проверка статистических гипотез.	Statistical hypotheses used in real case examples/ Types of sampling/ Calculating the error of sampling/ Restriction of Student's t-test Статистические гипотезы, используемые в реальных примерах. Типы отбора проб. Расчет погрешности выборки. Ограничение t-теста (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3)
4	Theme 4. Correlation and regression analysis Тема 4. Корреляционно-регрессионный анализ	Correlation and regression analysis in packages of applied statistical programs Корреляционный и регрессионный анализ в пакетах прикладных статистических программ (ПКос-8.1, ПКос-8.2, ПКос-8.3, ПКос-9.2, ПКос-9.3)
<b>Раздел 2. Интеллектуальный анализ данных</b>		
5	Тема 5. Понятие, сущность и задачи интеллектуального анализа данных	Fields of practical application of data mining. Methodology of data collection Basic types of databases Области практического применения интеллектуального анализа данных. Методология сбора данных Основные типы баз данных (УК -4.3, УК -4.3, ПКос-8.1)
6	Тема 6. Решение задач классификации и регрессии	One- and multidimensional classification Validation of predictive regression models Using the regression analysis in various applied fields

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
		Одно- и многомерная классификация Валидация моделей прогнозной регрессии Использование регрессионного анализа в различных прикладных областях (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3)
7	Тема 7. Поиск ассоциативных правил в процессе анализа данных	AIS-algorithm. SETM-algorithm. DIC-algorithm AIS –алгоритм. SETM-алгоритм. DIC-алгоритм (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3)
8	Тема 8. Кластерный анализ	Clustering and segmentation software. Measures of distances in cluster analysis. Sequence analysis Программное обеспечение для кластеризации и сегментации Меры расстояний в кластерном анализе Анализ последовательности (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3)
9	Тема 9. Нейросетевые методы анализа	Spheres of application of neural network analysis methods Сферы применения нейросетевых методов анализа (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3)
10	Тема 10. Анализ текстовой информации и аналитика	Text Mining in R. Statistical methods for text mining. Contextual text mining techniques. Текст Mining в R. Статистические методы для добычи текста. Методы добычи контекстного текста (УК -4.3, УК -4.3, ПКос-8.1)
11	Тема 11. Визуализация данных	Current trends in data visualization. Examples of interactive data visualization. Examples of static data visualization. 3D visualization. Современные тенденции в визуализации данных. Примеры интерактивной визуализации данных. Примеры визуализации статических данных. 3D визуализация (УК -4.3, УК -4.3, ПКос-8.1, ПКос-9.2, ПКос-9.3)

## 5. Образовательные технологии

Таблица 6

### Применение активных и интерактивных образовательных технологий

№ п/п	Тема и форма занятия		Наименование используемых активных и интерактивных образовательных технологий
1.	Классификация методов Data mining	Л	Лекция-визуализация
2.	Практическое занятие № 4. Разработка статистических тестов на языке R	ПЗ	Деловая игра

## 6. Текущий контроль успеваемости и промежуточная аттестация по итогам освоения дисциплины

## 6.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности

### 6.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности

#### Контрольная работа №1

##### Вариант 1

1. Создать вектор  $v$  (2,5,8,2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2) Создать вектор  $b$  из 20 случайных чисел в диапазоне от 1 до 100 без повторений.

Найти: для вектора  $b$ : среднюю (по формуле и по встроенной функции), среднеквадратическое отклонение (округлить в большую сторону), минимальное и максимальное значение, количество элементов.

Вектор  $v$  отсортировать в порядке возрастания. Найти значения элементов по ранжированному ряду с номерами 2 и 8.

Рассчитать ковариацию между векторами  $v$  и  $b$  ( $\text{cov} = \overline{xy} - \bar{x} * \bar{y}$ ).

2. Проверить по данным базы iris равенство средних по переменной Sepal.Length для сортов setosa и versicolor.  $c = t$  фактическое Сделать вывод.

3. Провести тест на независимость выборок для данных:

	A	B
C	30	15
D	10	13

4. По данным sleep построить регрессию зависимости лекарства (group) от прироста продолжительности сна (extra). Сделать вывод.

5. По данным mtcars построить зависимость топливной экономичности (mpg) от веса(wt) и вместимостью двигателя (disp). Сделать вывод.

#### Examination №1

##### Option 1

1. Create a vector  $v$  (2,5,8,2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2) Create vector  $b$  from 20 random numbers in the range from 1 to 100 without repetitions.

Find: for the vector  $b$ : the mean (by the formula and the built-in function), the mean-square deviation (rounded up), the minimum and maximum values, the number of elements.

Sort vector  $v$  in ascending order. Find the values of the elements in the ranked row with numbers 2 and 8.

Calculate the covariance between the vectors  $v$  and  $b$  ( $\text{cov} =$ ).

2. Check, according to the iris database, the equality of the averages with respect to the Sepal.Length variable for the setosa and versicolor varieties.  $c = t$  actual Make a conclusion.

3. Conduct a sample independence test for the data:

	A	B
C	30	15
D	10	13

4. According to sleep, construct a regression of drug dependence (group) on the increase in the duration of sleep (extra). Make a conclusion.

5. According to mtcars to build the dependence of fuel efficiency (mpg) on weight (wt) and engine capacity (disp). Make a conclusion.

### Вариант 2

1. Создать вектор  $v$  (2,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8)

Создать вектор  $b$  из 30 случайных чисел в диапазоне от -100 до 100 без повторов.

Найти: для вектора  $b$ : медиану (округлить до целого числа), дисперсию (округлить в большую сторону), сумму элементов, произведение элементов, количество элементов.

Вектор  $v$  отсортировать в порядке возрастания. Найти значения элементов по ранжированному ряду с номерами 4 и 15.

Рассчитать коэффициент корреляции между векторами  $v$  и  $b$  ( $r = \frac{xy - x \cdot y}{\sigma_x \cdot \sigma_y}$ ).

2. Сравнить на соответствие нормальному закону по критерию Шапиро-Уилка базы iris переменной Sepal.Length.  $h = W$

3. Проверить равенство средних для групп:

№ группы	$x_i$
1	2
1	8
1	6
1	5
1	7
2	4
2	3
2	5
2	6
2	5
3	3
3	4
3	3
3	6
3	5

4. Рассчитать несмещенную оценку дисперсии для №+10 переменных Sepal.Length сорта versicolor. № - номер Вашей команды.  $i = S^2$

5. По данным sleep построить регрессию зависимости лекарства (group) от прироста продолжительности сна (extra). Сделать вывод.

### Option 2.

1. Create a vector  $v$  (2,5,8,5,8,8,8,8,8,5,8,5,8,8,8,8,8,8,8,5,8,5,8,5, 8.5,8)

Create vector  $b$  from 30 random numbers in the range from -100 to 100 without repetitions.

Find: for vector b: median (rounded to a whole number), variance (rounded up), the sum of elements, the product of elements, the number of elements.

Sort vector v in ascending order. Find the values of the elements in the ranked row with numbers 4 and 15.

Calculate the correlation coefficient between the vectors v and b ( $r =$ ).

2. Compare for compliance with the normal law according to the Shapiro-Wilk criterion of the iris database Sepal.Length variable.  $h = w$

3. Check the equality of averages for groups:

№ group	$x_i$
1	2
1	8
1	6
1	5
1	7
2	4
2	3
2	5
2	6
2	5
3	3
3	4
3	3
3	6
3	5

4. Calculate the unbiased estimate of the variance for the No. + 10 Se-pal.Length variable varieties of versicolor. No. is your team number.  $i = S^2$

5. According to sleep, construct a regression of drug dependence (group) on the increase in the duration of sleep (extra). Make a conclusion.

## Контрольная работа № 2

### Вариант 1

По данным о студентах предсказать поступит он или нет методом «случайный лес» и методом опорных векторов.

### Вариант 2 .

По данным о студентах предсказать поступит он или нет методом опорных векторов с ядерной функцией и методом «дерево решений»

### Вариант 3



По данным о студентах предсказать поступит он или нет с использованием нейронных сетей и с помощью Байесовской классификации.

## **Examination number 2**

### **Option 1**

According to the data about the students, it is predicted whether or not he will use the random forest method and the support vector method.

### **Option 2.**

According to the data about the students, whether or not to predict whether it will come using the method of supporting vectors with the nuclear function and the “decision tree” method

### **Option 3**

According to the data on students, it is predicted whether or not he will act using neural networks and using Bayesian classification.

#### **1. What is data mining?**

- a) extraction of information from raw data
- b) storing and processing the information in databases
- c) data visualization
- d) statistical methods of data processing

#### **2. What concept is the basis of data mining technology?**

- a) the concept of natural science
- b) the concept of management
- c) the concept of patterns
- d) the concept of formation

#### **3. What is the result of representing the obtain knowledge in data mining?**

- a) pattern
- b) project
- c) model
- d) trend

#### **4. The task of classification is a \_\_\_\_\_ task**

- a) descriptive
- b) predictive
- c) qualitative
- d) quantitative

#### **5. The task of clustering is a \_\_\_\_\_ task**

- a) descriptive
- b) predictive
- c) qualitative
- d) quantitative

**6. Data objects with characteristics that are considerably different than most of the other data objects in the data set:**

- a) OLTP
- b) ordinal
- c) interval
- d) outliers

**7. A set of examples with a known outcome:**

- a) clients
- b) classifiers
- c) training set
- d) data mining

**8. Data transformation includes...**

- a) separating data from one source into various sources of data
- b) a process to change data from a summary level to a detailed level
- c) joining data from one source into various sources of data
- d) a process to change data from a detailed level to a summary level

**9. The \_\_\_\_\_ is a useful method of discovering patterns at the beginning of data mining process.**

- a) calculating distance
- b) decision trees
- c) association rules
- d) visualization techniques

**10. \_\_\_\_\_ is a class of models based on biological analogy with a human brain:**

- a) neural network
- b) cluster
- c) decision tree
- d) rule of classification

**11. What kind of hierarchical structure decision trees create?**

- a) if... then...
- b) neither... nor...
- c) either... or...
- d) both... and...

**12. Which of the following is the mathematical equation connecting the variables x and y?**

- a) regression
- b) interpolation
- c) slope b
- d) extrapolation

**13. The equation of the form  $y = a + bx$ :**

- a) y-intercept a
- b) linear regression
- c) regression
- d) interpolation

**14. Which of these indicators is a measure of the fit of a regression model?**

- a) mean
- b) variance
- c) standard deviation
- d) coefficient of determination

**15. Coefficient of correlation always lies between...**

- a) 0 and 1
- b) -1 and 1
- c) -1 and 0
- d) 0 and 2

**16. Investigating the relationship between several variables is the task of...**

- a) simple regression
- b) multiple regression
- c) decision trees
- d) modeling

**17. Which of the following is a method of constructing classification rules?**

- a) 1R-algorithm
- b) Naive Bayes method
- c) both
- d) none

**18. The expected mean value of y when  $X = 0$**

- a) constant
- b) intercept
- c) R
- d) R squared

**19. Series of data points indexed in time order is a...**

- a) panel data
- b) cross-sectional data
- c) time series data
- d) none

**20. It is possible to investigate the nature of the relationships between two variables by...**

- a) XY-plots
- b) Histograms

- c) Time series graphs
- d) none

**21. The main aim of association rule mining is to...**

- a) construct the rules of classification
- b) validate the predictive regression model
- c) define the main idea of the text
- d) discover patterns between related events in databases

**22. If several events are connected to each other, then it is...**

- a) association
- b) sequence
- c) classification
- d) clustering

**23. The ratio of transactions which have the set F ( $D_F$ ) to the whole number of transactions (D) is called...**

- a) the level of the set F
- b) the model of the set F
- c) the cluster of the set F
- d) the support of the set F

**24. The set of items (F) is large when...**

- a)  $\text{Supp}(F) < \text{Supp}(\text{min})$
- b)  $\text{Supp}(F) > \text{Supp}(\text{min})$
- c)  $\text{Supp}(F) = \text{Supp}(\text{min})$
- d) none

**25. The objects of some item set that are subjected to analysis are called...**

- a) ratios
- b) centroids
- c) transactions
- d) models

**26. What is a sequence in association rule mining?**

- a) finite real numbers
- b) range
- c) ordered set of some objects
- d) none

**27. Which of the following is not an evaluation of utility of association rules?**

- a) quality
- b) support
- c) confidence
- d) improvement

- 28. The support of any item set can't be more than the minimum support of any of its subsets. It is the main property of...**
- a) Naive Bayes method
  - b) Apriori algorithm
  - c) decision tree
  - d) clustering
- 29. The one of the methods of cluster analysis is...**
- a) web mining
  - b) k-means
  - c) regression
  - d) variance
- 30. Which of algorithms is hierarchical?**
- a) agglomerative
  - b) k-means
  - c) Fuzzy C-Means
  - d) none
- 31. Which of the following is not a measure of distances between clusters?**
- a) nearest neighbor
  - b) furthest neighbor
  - c) median clustering
  - d) mode clustering
- 32. Which of algorithms is non-hierarchical?**
- a) agglomerative
  - b) divisive
  - c) k-means
  - d) none
- 33. What are clusters?**
- a) time series data
  - b) homogenous groups of objects
  - c) heterogenous groups of objects
  - d) statistical population
- 34. One of the measure of similarity between pairs of objects is...**
- a) Chebyshev distance
  - b) Euclidian distance
  - c) both
  - d) none
- 35. A common way to visualize the cluster analysis's progress is by drawing...**
- a) a histogram
  - b) XY-plot

- c) dendrogram
- d) time series graph

**36. One of the criteria of an evaluation choice for a clustering solution is substantial. It means...**

- a) Segments should be relevant in respect of the company's competencies and objectives
- b) The segments are large and profitable enough to serve
- c) Effective programs can be formulated to attract and serve the segments
- d) The segments can be effectively reached and served, which requires them to be characterized by means of observable variables

**45. A graph can be convenient powerful representation for data if...**

- a) graph captures relationships among data objects
- b) the data objects are represented as graphs
- c) both
- d) none

**46. The first phase of data visualization is...**

- a) a cursory review
- b) a constructing the graph
- c) an interpreting results
- d) an evaluation of effectiveness

**47. One of the methods of visualization is...**

- a) clustering
- b) text mining
- c) geometric transformations
- d) none

**48. Which of the following is not graphic type?**

- a) bar chart
- b) histogram
- c) range
- d) XY-plot

**49. Data visualization allows us to spot...**

- a) patterns
- b) trends
- c) correlations
- d) all

**50. Which of these indicators is a measure of the fit of a regression model?**

- a) mean
- b) variance
- c) standard deviation

d) coefficient of determination

**51. At the beginning of the statistical research we have two statistical hypotheses:**

- a) null and alternative
- b) univariate and bivariate
- c) correct and incorrect
- d) none

**52. Statistical hypothesis - is:**

- a) any assumption used in statistical research;
- b) the assumption that can be tested using the available statistical information;
- c) scientific hypothesis put forward to explain a phenomenon and should be checked by experiment.

**53. Criterion - is:**

- a) hallmark received as the norm, the measure
- b) that certifies the knowledge of objective truth
- c) a set of rules adopted for the significance test

**54. Power of the test is:**

- a) objects introduced into the production process
- b) the ability to clearly distinguish criterion null and alternative hypotheses statistics
- c) the value of which is determined by the amount of power developed by the engine.

**55. Error of the first series - is:**

- a) the adoption of statistical hypothesis when it is wrong
- b) Deviation statistical hypothesis when it is correct
- c) error in determining the true value of the characteristic
- d) error in calculating the statistic

**56. Error of the second series - is:**

- a) the adoption of statistical hypothesis when it is wrong
- b) Deviation statistical hypothesis when it is correct
- c) error in determining the true value of the characteristic
- d) error in calculating the statistic

**57. Level of significance - is:**

- a) the probability with which guarantees reliable calculation of an indicator
- b) the value of quantitative indicator of the degree of manifestation or quality indicator
- c) the probability corresponding deflection correct hypothesis.

**58. Critical range of values - is:**

- a) the maximum values of feature;
- b) the minimum characteristic value;

c) area, getting value of the statistical criterion, which leads to the rejection of the test of statistical hypotheses.

### **Примеры текстов для чтения и перевода**

#### **1. From statistic... to data mining**

Traditional statistics (up to the 1950s):

- • A few hundred individuals
- • Several variables defined with a special protocol (sampling, experimental design, etc)
- • Firm assumptions regarding the statistical distributions involved (linearity, normality and homoscedasticity)
- • Models developed theoretically and compared with the data
- • Probabilistic and statistical methods
- • Used in the laboratory.

Data analysis (1960-1980):

- • A few thousand individuals
- • Several tens of variables
- • Construction “individuals x variables” tables
- • Importance of computing and visual representation.
- 

Data mining (1990s onwards):

- • Several millions or tens of millions of individuals
- • Several hundreds or thousands of variables
- • Numerous non-numeric variables, such as textual variables (or variables containing images)
- • Weak assumptions regarding the statistical distributions involved
- • Data collected before the study, and often for other purposes
- • Constantly changing population (difficulty of sampling)
- • Presence of “outliers” (abnormal individuals, at least in terms of the distributions studied)
- • Imperfect data, with errors of input and coding, and missing values
- • Fast computing, possibly in real time, is essential
- • The aim is not always to find the mathematical optimum, but sometimes the model that is easiest for non-statisticians to understand
- • The models are developed from the data, and attempts are sometimes made to draw theoretical conclusions from them
- • Use of statistical methods, artificial intelligence and machine learning theory
- • Used in the business world.

#### **2. The definition of data mining and its tasks**



- Data mining is the set of methods and techniques for exploring and analyzing data sets (which are often large), in an automatic or semi-automatic way in order to find among these data certain unknown or hidden rules, associations or tendencies.
- Briefly, data mining is the art of extracting information – that is, knowledge – from data.
- Methods of data mining enable us to solve a lot of tasks. The main of them are classification task, regression task, association rule mining and clustering.

The classification task is to determine the class of object by its characteristics. The regression task helps us to determine the value of some object parameter. This value is the set of all real numbers. Association rule mining aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in large databases. The clustering task is to find independent groups (or clusters) and their characteristics among the set of analyzed data. Grouping of homogeneous objects enables us to reduce their number and make analysis much easier.

- Listed tasks are divided into descriptive and predictive tasks. The descriptive techniques are designed to bring out information that is present but buried in a mass of data, while the predictive techniques are designed to extrapolate new information based on the present information, this new information being qualitative (in the form of classification or scoring) or quantitative (regression).

### **3. The practical use of data mining**

- The main fields where data mining is used: banking (12%), customer relationship management (12%), direct marketing (8%), fraud detection (7%), insurance (6%), retail (6%), telecommunications (5%), scientific research (4%), and health (4%). It is based on a survey on the [www.kdnuggets.com](http://www.kdnuggets.com) portal.

Data mining for banking sector:

- • it is essential for efficient marketing, which relies on the size and potential of the customers;
- • it allows to tackle accurately customer needs;
- • the commercial banking market is lucrative and offers challenges that can be solved with data mining.
- In retail industry data mining can help identify customer behavior, discover customer shopping patterns and trends, improve the quality of customer service, achieve better customer retention and satisfaction, enhance goods consumption ratios, design more effective goods transportation and distribution policies and reduce the cost of business.
- The insurance industry is totally dependent on the ability to convert raw data into intelligence – intelligence about customers, markets, competitors, and business environment.

- In medical sector data mining researches such major areas as the evaluation of treatment effectiveness, management of healthcare, customer relationship management, and the detection of fraud and abuse.
- The food industry is also a major user of data mining. Its application includes «sensory analysis» in which sensory data (taste, flavor, consistency, etc.) perceived by consumers are correlated with physical and chemical instrumental measurements and with preference for various products.

#### The R environment

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. Among other things it has

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either directly at the computer or on hardcopy, and
- a well developed, simple and effective programming language (called ‘S’) which includes conditionals, loops, user defined recursive functions and input and output facilities. (Indeed most of the system supplied functions are themselves written in the S language.)

The term “environment” is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.

R is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly, and has been extended by a large collection of packages. However, most programs written in R are essentially ephemeral, written for a single piece of data analysis.

### Примерные темы для обсуждения

#### Практическая работа № 1 Предмет и метод и задачи статистики

1. Определение понятия «статистика».
2. Чем вызвано появление и развитие науки и практики статистики?
3. Какие явления изучает статистика, что является объектом ее исследования?
4. Охарактеризуйте понятие «признак» и опишите его виды.
5. Чем различаются понятия «показатель» и «статистический показатель»?
6. Укажите отличительные особенности статистических показателей.
7. Что является предметом статистики?
8. Раскройте содержание системы статистических показателей.
9. В чем состоит сущность статистических закономерностей?
10. Охарактеризуйте основные черты метода статистики и этапы статистического исследования.
11. Как организована статистика в мире и в современной России?

## **Practical work No. 1 Subject and method and tasks of statistics**

1. The definition of "statistics".
2. What caused the emergence and development of the science and practice of statistics?
3. What phenomena is studied by statistics, what is the object of its research?
4. Describe the concept of "sign" and describe its types.
5. What is the difference between the concept of "indicator" and "statistical indicator"?
6. Indicate the distinctive features of statistical indicators.
7. What is the subject of statistics?
8. Expand the content of the system of statistical indicators.
9. What is the essence of statistical laws?
10. Describe the main features of the statistical method and the stages of statistical research.
11. How is the statistics organized in the world and in modern Russia?

### **Практическое занятие №6 «Интеллектуальный анализ данных и статистика: в чём различия?»**

1. Понятие интеллектуального анализа данных.
2. Основные этапы развития интеллектуального анализа данных.
3. Основные задачи интеллектуального анализа данных.
4. Описательные и предсказательные задачи.
5. Какие статистические методы могут применяться в Data Mining.
6. Основные области практического применения интеллектуального анализа данных.
7. Как интеллектуальный анализ данных используется экономистами?

### **Practical lesson № 6 “Intellectual data analysis and statistics: what are the differences?”**

1. The concept of data mining.
2. The main stages of the development of data mining.
3. The main tasks of data mining.
4. Descriptive and predictive tasks.
5. What statistical methods can be used in Data Mining.
6. The main areas of practical application of data mining.
7. How is data mining used by economists?

### **Практическое занятие №20«Анализ текстовой информации: подходы, особенности, перспективы»**

1. Основные этапы текстового анализа.
2. Задачи текстового анализа.
3. Извлечение ключевых понятий из текста.
4. Классификация документов.
5. Кластерный анализ документов.
6. Существующие программные обеспечения в области анализа текстовой информации.

### **Practical lesson №20 "Analysis of textual information: approaches, features, prospects"**

1. The main stages of text analysis.

2. Tasks of text analysis.
3. Extraction of key concepts from the text.
4. Classification of documents.
5. Cluster analysis of documents.
6. Existing software in the field of text information analysis.

**Практическое занятие №22 «Классификация методов визуализации данных»**

1. Характеристики инструментов визуализации данных.
2. Методы визуализации.
3. Существующие тренды в области визуализации данных.
4. 3D визуализация.

**Practical lesson №22 "Classification of data visualization methods"**

1. Characteristics of data visualization tools.
2. Visualization methods.
3. Existing trends in data visualization.
4. 3D visualization.

**Практическое занятие №23 «Программное обеспечение для визуализации данных»**

1. OfficeReports
2. Second Prism
3. Databoard
4. DataMarket
5. Q Research Software
6. Statwing

**Practical lesson №23 "Software for data visualization"**

1. OfficeReports
2. Second Prism
3. Databoard
4. DataMarket
5. Q Research Software
6. Statwing

**Пример работ**

**Практическое занятие №2 «Расчет основных показателей центральной тенденции и вариации в пакете R»**

Загрузить исходные данные в RStudio. Написать функцию, для расчета среднего значения, моды и медианы, дисперсии, размаха вариации и коэффициента вариации для ранжированного, дискретного и интервального рядов распределения. Рассчитать эти показатели с использованием встроенных функций. Сравнить значения. Оформить отчет с выводами.

**Practical lesson number 2 "Calculation of the main indicators of the central trends and variations in the package R"**

Download the source data in RStudio. Write a function to calculate the average value, mode and median, variance, range of variation and coefficient of variation for the ranked, discrete and interval series of the distribution. Calculate these indicators using built-in functions. Compare values. To issue a report with the findings.

### **Практическое занятие № 3. Статистические тесты в среде R**

Загрузить исходные данные в RStudio.

1. Проверить соответствие исходных данных нормальному закону распределения с использованием критериев Колмогорова-Смирнова, Уилкоксона, Шапиро-Уилка, Пирсона.

2. Сравнить различия в средних по критерию Стьюдента, F-Фишера.

3. Построить таблицу частот по двум признакам. Проверить независимость выборок по критерию  $\chi^2$  – Пирсона.

Оформить отчет с выводами.

#### **Practical lesson № 3. Statistical tests in the environment of R**

Download the source data in RStudio.

1. Check the compliance of the source data with the normal distribution law using the Kolmogorov-Smirnov, Wilcoxon, Shapiro-Wilk, Pearson criteria.

2. To compare differences in averages by criterion of Student, F-Fisher.

3. Build a table of frequencies in two ways. Check the independence of the samples according to the  $\chi^2$  - Pearson criterion.

To issue a report with the findings.

### **Практическое занятие № 4. Разработка статистических тестов на языке R**

По исходным данным предыдущей задачи написать функция для применения дисперсионного анализа, критерия t-Стьюдента. Для дисперсионного анализа предусмотреть процедуру попарного сравнения методом Q-Тьюки. Оформить отчет с выводами.

#### **Practical lesson № 4. Development of statistical tests in the language of R**

According to the initial data of the previous task, write a function for applying the analysis of variance, the t-Student criterion. For analysis of variance, provide a pair-wise comparison procedure using the Q-Tukey method. To issue a report with the findings.

### **Практическое занятие № 5. Корреляционно-регрессионный анализ в среде R**

Загрузить исходные данные в RStudio. Построить уравнение парной линейной регрессии. Сделать прогноз по уравнению. Построить множественное уравнение линейной регрессии. Сделать прогноз по уравнению. Вывести summary. Оформить отчет с выводами.

#### **Practical lesson number 5. Correlation and regression analysis in the environment of R**

Download the source data in RStudio. Build the equation of pairwise linear regression. Make a prediction by equation. Build a multiple linear regression equation. Make a prediction by equation. Print summary. To issue a report with the findings.

#### **Вопросы к зачету**

1. General definition of statistics
2. The subject of statistics
3. The method of statistics
4. The main concepts of statistics
5. Differences between data mining and statistics

6. Types of measurements units used for measurement of absolute statistical indicators.
7. Types of relative values. Their calculation procedure.
8. What is a mean value? The exponential means, and the procedure used for their calculation.
9. Structural means and the procedure for their calculation.
10. What is the system of statistical indicators?
11. The main types of statistical testing hypotheses.
12. Level of significance criterion.
13. The common method of constructing confidence intervals and statistical hypothesis testing.
14. The meaning of the concepts of "error of the first kind", "error of the second kind", "power of the test".
15. Critical limits for single and double criteria for a given level of significance.
16. Give examples of practical problems in testing hypotheses about the equality of the expectation, variance.

### **Вопросы к экзамену**

1. The definition of data mining.
2. The main phases of data mining development.
3. The main tasks of data mining.
4. Descriptive and predictive tasks.
5. Regression and classification tasks.
6. Association rule mining task.
7. Clustering task.
8. The main fields of practical application of data mining.
9. How data mining is used by economists.
10. Defining the aim of data mining project.
11. Collecting the data for implementation of data mining project.
12. Database creation for storing and processing data.
13. Exploring and preparing the data.
14. Drawing up and validating the predictive models.
15. Deploying the models.
16. Training the models users.
17. Monitoring and enriching the models.
18. Ways of representing relationships between variables.
19. Rules of classification.
20. Decision trees.
21. Mathematical functions.
22. 1R-algorithm.
23. Naive Bayes method.
24. ID3 and C4.5 algorithms for constructing the decision trees.
25. Linear methods of constructing the equations of describing relationships between variables.

26. Ordinary least squares (OLS).
27. Nonlinear methods.
28. Problem statement of association rule mining.
29. Sequence analysis.
30. Representing the results of association rule mining.
31. Apriori algorithm.
32. Problem statement of clustering.
33. Measures of distances in cluster analysis: Euclidean distance, Chebyshev distance.
34. Representing the results of cluster analysis.
35. Classification of clustering algorithms.
36. Hierarchical algorithms: agglomerative and divisive methods of clustering.
37. Non-hierarchical algorithms: k-means, Fuzzy C-Means.
38. The main phases of text analysis.
39. Tasks of text mining.
40. Extracting central concepts from text.
41. Classification of text documents.
42. Clustering of text documents.
43. Text mining software packages.
44. Characteristics of tools of data visualization.
45. Methods of visualization.
46. The main trends in data visualization.
47. 3D visualization.

## **6.2. Описание показателей и критериев контроля успеваемости, описание шкал оценивания**

Текущий контроль знаний, умений и навыков проводится в форме тестирования и контрольных работ с расчетными задачами и теоретическими вопросами, а также предусмотрена защита индивидуальных задач, выполняемых каждым студентом на практических занятиях. Ликвидация студентами текущих задолженностей производится также в форме выполнения индивидуальной задачи по соответствующей теме и дальнейшей ее защиты преподавателю кафедры.

Итоговая оценка учитывает результаты рейтинговой системы контроля знаний (вклад 50%), результаты контрольных работ по всем темам (вклад 50%) в VII семестре, результаты ответа на экзаменационные вопросы в VIII семестре.

Вид промежуточного контроля по данному направлению: VII семестр – зачет; VIII семестр – экзамен. Критерии выставления оценок по системе:

0-59 % от максимального количества баллов – «незачтено»;

60 и более %– «зачтено».

Критерии выставления оценок на экзамене по четырехбалльной системе:

Неудовлетворительно – 0-59% от максимального количества баллов;

Удовлетворительно – 60 – 74%;

Хорошо – 75 – 84%;

Отлично – 85 – 100%.

## 7. Учебно-методическое и информационное обеспечение дисциплины

### 7.1 Основная литература

1. Макшанов, А. В. Технологии интеллектуального анализа данных : учебное пособие / А. В. Макшанов, А. Е. Журавлев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 212 с. — ISBN 978-5-8114-4493-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/206711> (дата обращения: 14.12.2022). — Режим доступа: для авториз. пользователей.
2. Остроух, А. В. Системы искусственного интеллекта : монография / А. В. Остроух, Н. Е. Суркова. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 228 с. — ISBN 978-5-8114-8519-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/176662> (дата обращения: 14.12.2022). — Режим доступа: для авториз. пользователей.
3. Митина, О. А. Языки программирования для статистической обработки данных (R) : учебное пособие / О. А. Митина. — Москва : РТУ МИРЭА, 2020. — 191 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/163912> (дата обращения: 15.11.2022). — Режим доступа: для авториз. пользователей.
4. Алексеев, Д. С. Технологии интеллектуального анализа данных : учебник для вузов / Д. С. Алексеев, О. В. Щекочихин. — Санкт-Петербург : Лань, 2022. — 176 с. — ISBN 978-5-8114-8299-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/187559> (дата обращения: 14.12.2022). — Режим доступа: для авториз. пользователей.
5. Демидова, Л. А. Интеллектуальный анализ данных на языке Python : учебно-методическое пособие / Л. А. Демидова. — Москва : РТУ МИРЭА, 2021. — 92 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/218693> (дата обращения: 14.12.2022). — Режим доступа: для авториз. пользователей.

### 7.2 Дополнительная литература

1. Перекаатов, А. С. Статистическая обработка экспериментальных данных. Полный факторный эксперимент в языке R : учебное пособие / А. С. Перекаатов, М. Б. Никифоров. — Рязань : РГРТУ, 2019. — 48 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/168309> (дата обращения: 28.11.2022). — Режим доступа: для авториз. пользователей.
2. Гришин, В. А. Методы обработки данных и моделирование на языке R : учебно-методическое пособие / В. А. Гришин, М. С. Тихов. — Нижний Новгород : ННГУ им. Н. И. Лобачевского, 2019. — 54 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL:



<https://e.lanbook.com/book/144653> (дата обращения: 15.11.2022). — Режим доступа: для авториз. пользователей.

3.

4. Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python : учебное пособие / В. М. Волкова, М. А. Семёнова, Е. С. Четвертакова, С. С. Вожов. — Новосибирск : НГТУ, 2017. — 74 с. — ISBN 978-5-7782-3183-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/118287> (дата обращения: 15.11.2022). — Режим доступа: для авториз. Пользователей.

5. Гильванов, Р. Г. Технологии обработки информации: Лабораторный практикум : учебное пособие / Р. Г. Гильванов, , А. В. Забродин. — Санкт-Петербург : ПГУПС, 2022. — 45 с. — ISBN 978-5-7641-1819-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/279065> (дата обращения: 28.11.2022). — Режим доступа: для авториз. пользователей.

6. Зыков, С. В. Программирование : учебник и практикум для вузов / С. В. Зыков. — Москва : Издательство Юрайт, 2022. — 320 с. — (Высшее образование). — ISBN 978-5-534-02444-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489754>

### **7.3 Методические указания, рекомендации и другие материалы к занятиям**

1. Харитонов, А.Е. Хранилища и системы интеллектуального анализа данных: методические указания / А.Е. Харитонов. — М.: РГАУ-МСХА им. К.А. Тимирязева, 2016. — 25 с.

### **8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины**

1. The R Project for Statistical Computing <https://www.r-project.org/> (открытый доступ)

2. The R Project for Statistical Computing <https://www.r-project.org/> (открытый доступ)

3. Анаконда. URL: <https://www.anaconda.com/distribution/> (открытый доступ)

4. Официальный сайт Росстата. URL: <http://www.gks.ru/> (открытый доступ)

5. Официальный сайт Центрального Банка России. URL: <http://www.cbr.ru> (открытый доступ)

6. Bureau of Economic Analysis. URL: <http://www.bea.gov> (открытый доступ)

7. Московская международная валютная биржа. <http://www.micex.ru> (открытый доступ)

8. Официальный сайт Всемирного банка . URL: <http://www.worldbank.org> (открытый доступ)

9. Официальный сайт Министерства финансов РФ. URL: <http://www.minfin.gov.ru> (открытый доступ)

10.Официальный сайт Национального бюро экономических исследований США. URL: [http:// www.nber.org](http://www.nber.org) (открытый доступ)

## 9. Перечень программного обеспечения и информационных справочных систем

Таблица 9

### Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	1-11	R	расчётная	r-project	2022
2	1-11	RStudio	расчётная	r-project	2022
3	1-11	Anaconda	расчётная	Anaconda Enterprise	2022

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

Таблица 10

### Сведения об обеспеченности специализированными аудиториями, кабинетами, лабораториями

Наименование специальных помещений и помещений для самостоятельной работы (№ учебного корпуса, № аудитории)	Оснащенность специальных помещений и помещений для самостоятельной работы
1	2
<i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации</i>	<ol style="list-style-type: none"> <li>1. Экран с электроприводом 1 шт. (Инв. №558771/2)</li> <li>2. Проектор 1 шт. (без инв. №) – приобретался не за счет средств вуза</li> <li>3. Вандалоустойчивый шкаф 1 шт. (Инв.№558850/7)</li> <li>4. Системный блок с монитором 1 шт. (Инв. №558777/9)</li> <li>5. Стенд «Сергеев Сергей Степанович 1910-1999» 1 шт. (Инв.№591013/25)</li> <li>6. Огнетушитель порошковый 1 шт. (Инв. №559527)</li> <li>7. Подвесное крепление к огнетушителю 1 шт. (Инв. № 559528)</li> <li>8. Жалюзи 2шт. (Инв. №1107-221225, Инв. №1107-221225)</li> <li>9. Лавка 20 шт.</li> <li>10. Стол аудиторный 20 шт.</li> <li>11. Стол для преподавателя 1 шт.</li> <li>12. Стул 2 шт.</li> <li>13. Доска маркерная 1 шт.</li> <li>14. Трибуна напольная 1 шт. (без инв. №)</li> </ol>
<i>учебная аудитория для проведения занятий семинарского типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации, помещение для самостоятельной работы</i>	<ol style="list-style-type: none"> <li>1. Системный блок Intel Core Intel Core i3-2100/4096Mb/500Gb/DVD-RW 10 шт. (Инв.№601997, Инв.№601998, Инв.№601999, Инв.№602000, Инв.№602001, Инв.№602002, Инв.№602003, Инв.№602004, Инв.№602005, Инв.№602006)</li> <li>2. Монитор 10 шт. (без инв. №) - приобретались не за счет средств вуза</li> <li>3. Шкаф 2 шт. (Инв.№594166, Инв.№594167)</li> <li>4. Тумба 1 шт. (Инв.№594168)</li> <li>5. Подвесное крепление к огнетушителю 1 шт. (Инв. № 559528)</li> <li>6. Огнетушитель порошковый 1 шт. (Инв. №559527)</li> <li>7. Жалюзи 1 шт. (Инв.№551557)</li> </ol>

	8. Доска магнитно-маркерная 1 шт. 9. Стол 5 шт. 10. Стол компьютерный 12 шт. 11. Стул офисный 21 шт. 12. Сейф 1 шт. (без Инв.№).
<i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации, помещение для самостоятельной работы</i>	1. Трибуна напольная 1 шт. (Инв.№ 599205) 2. Шкаф для документов 3 шт. (Инв.№593633, Инв.№593634, Инв.№559548/18) 3. Вешалка напольная 2 шт. (Инв.№1107-333144, Инв.№1107-333144) 4. Жалюзи 1 шт. (Инв.№591110) 5. Доска магнитно-маркерная 1 шт. 6. Стол 15 шт. 7. Скамейка 14 шт. 8. Стол эрго 1 шт. 9. Стул 2 шт.
<i>Центральная научная библиотека имени Н.И. Железнова</i>	Читальные залы библиотеки
<i>Студенческое общежитие</i>	Комната для самоподготовки

## 11. Методические рекомендации обучающимся по освоению дисциплины

Приступая к изучению дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке», студенты должны ознакомиться с учебной программой, учебной, научной и методической литературой, имеющейся в библиотеке РГАУ-МСХА им. К.А. Тимирязева, получить в библиотеке рекомендованные учебники и учебно-методические пособия, завести новую тетрадь для работы с первоисточниками.

В ходе занятий вести конспектирование учебного материала, обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации, записывать слова и словосочетания на иностранном языке по теме занятия, задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций или касательно иностранных слов, вызывающих затруднения при переводе.

В ходе подготовки к практическим занятиям изучить основную литературу, ознакомиться с дополнительной литературой в соответствии с поставленной задачей. При этом учесть рекомендации преподавателя и требования учебной программы. Необходимо дорабатывать свой конспект, делая в нем соответствующие записи из литературы, рекомендованной преподавателем и предусмотренной учебной программой.

При подготовке к зачету и экзамену повторять пройденный материал в строгом соответствии с учебной программой. Использовать конспекты и литературу, рекомендованную преподавателем. Обратит особое внимание на темы учебных занятий, пропущенных студентом по разным причинам. При необходимости обратиться за консультацией и методической помощью к преподавателю.

### Виды и формы отработки пропущенных занятий

Студент, пропустивший занятия обязан самостоятельно выполнить сообщение (презентацию), рассмотренную на практическом занятии и подготовиться по контрольным вопросам к защите работы в рамках часов консультаций.

## **12. Методические рекомендации преподавателям по организации обучения по дисциплине**

Курс интеллектуального анализа данных и статистики на иностранном языке должен давать не абстрактно-формальные, а прикладные знания. Данная цель может быть реализована только при условии соблюдения в учебных планах преемственности учебных дисциплин. Базовые знания для изучения методов интеллектуального анализа данных дают такие дисциплины, как иностранный язык, экономическая теория, информатика. Освоение подходов к интеллектуальному анализу данных и статистике позволит студентам в будущем использовать его в исследованиях, в планировании и прогнозировании.

Студент может провести собственное исследование на примере конкретного предприятия, представляющим его научный интерес, представить результаты в виде презентации. В случае надлежащего качества, его работа может быть заслушана на научном кружке кафедры или на студенческой научной конференции. По решению кафедры, студенты, занявшие призовые места на научных студенческих конференциях, могут освободиться от сдачи зачета или экзамена по этой дисциплине.

**Программу разработал (и):**

Харитонов А.Е., к.э.н., доцент



(подпись)

## РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.ДВ.06.01 «Интеллектуальный анализ данных и статистика на иностранном языке»

ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»  
(квалификация выпускника – бакалавр)

Коломеева Елена Сергеевна, доцент кафедры финансов ФГБОУ ВО г. Москвы «РГАУ-МСХА имени К.А. Тимирязева», кандидатом экономических наук (далее по тексту рецензент), проведена рецензия рабочей программы дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)» (бакалавриат) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре статистики и кибернетики (разработчик – Харитоновна Анна Евгеньевна, кандидат экономических наук, доцент кафедры статистики и кибернетики).

Рассмотрев представленные на рецензирование материалы, рецензент пришел к следующим выводам:

1. Предъявленная рабочая программа дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» (далее по тексту Программа) соответствует требованиям ФГОС ВО по направлению 09.03.02 «Информационные системы и технологии». Программа содержит все основные разделы, соответствует требованиям к нормативно-методическим документам.

1. Представленная в Программе **актуальность** учебной дисциплины в рамках реализации ОПОП ВО не подлежит сомнению – дисциплина относится к дисциплинам по выбору части, формируемой участниками образовательных отношений учебного цикла – Б1.В.ДВ.

2. Представленные в Программе **цели** дисциплины соответствуют требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

3. В соответствии с Программой за дисциплиной «Интеллектуальный анализ данных и статистика на иностранном языке» закреплено **3 компетенции (7 индикаторов)**. Дисциплина «Интеллектуальный анализ данных и статистика на иностранном языке» и представленная Программа способна реализовать их в объявленных требованиях.

4. Общая трудоёмкость дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» составляет 5 зачётных единиц (18 часов/из них практическая подготовка 8 ч.).

5. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Интеллектуальный анализ данных и статистика на иностранном языке» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии» и возможность дублирования в содержании отсутствует.

6. Представленная Программа предполагает использование современных образовательных технологий, используемые при реализации различных видов учебной работы. Формы образовательных технологий соответствуют специфике дисциплины.

7. Программа дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» предполагает занятия в интерактивной форме.

8. Виды, содержание и трудоёмкость самостоятельной работы студентов, представленные в Программе, соответствуют требованиям к подготовке выпускников, содержащимся во ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

9. Представленные и описанные в Программе формы *текущей* оценки знаний (опрос, как в форме обсуждения отдельных вопросов, так и участие в деловых играх), соответствуют специфике дисциплины и требованиям к выпускникам.

Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме зачета, что соответствует статусу дисциплины, как факультативной дисциплины учебного цикла – ФТД ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

10. Формы оценки знаний, представленные в Программе, соответствуют специфике дисциплины и требованиям к выпускникам.

11. Учебно-методическое обеспечение дисциплины представлено: основной литературой – 5 источников (базовый учебник), дополнительной литературой – 5 наименований, Интернет-ресурсы – 10 источников и соответствует требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

12. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» и обеспечивает использование современных образовательных, в том числе интерактивных методов обучения.

13. Методические рекомендации студентам и методические рекомендации преподавателям по организации обучения по дисциплине дают представление о специфике обучения по дисциплине «Интеллектуальный анализ данных и статистика на иностранном языке».

### ОБЩИЕ ВЫВОДЫ

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Интеллектуальный анализ данных и статистика на иностранном языке» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность **«Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»** (квалификация выпускника – бакалавр), разработанная Харитоновой А. Е., к.э.н., доцентом кафедры статистики и кибернетики, соответствует требованиям ФГОС ВО, современным требованиям экономики, рынка труда и позволит при её реализации успешно обеспечить формирование заявленных компетенций.

Рецензент: Коломеева Е.С., доцент кафедры финансов ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», кандидат экономических наук \_\_\_\_\_ « 26 » \_\_\_\_\_ 2022 г.  
(подпись)