

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Астраханский государственный университет имени В. Н. Татищева»  
(Астраханский государственный университет им. В. Н. Татищева)

СОГЛАСОВАНО  
Руководители ОПОП

А.Н. Бармин  
«4» июня 2024 г.

УТВЕРЖДАЮ

Заведующий кафедрой биотехнологии,  
аквакультуры, почвоведения и управления  
земельными ресурсами

Л.В. Яковлева  
«4» июня 2024 г.

Рабочая программа дисциплины (модуля)  
**«Основы BigData и анализа данных»**

Составитель

Сорокин А.П., доцент, к.б.н., доцент  
кафедры биотехнологии, аквакультуры, почво-  
ведения и управления земельными ресурсами

Направление подготовки /  
специальность

**05.03.06 Экология и природопользование**

Направленность (профиль) ОПОП

**Геоэкология**

Квалификация (степень)

**Бакалавр**

Форма обучения

**Очная**

Год приёма

**2023**

Курс

**2**

Семестр

**3**

## **1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)**

**1.1. Целями освоения дисциплины (модуля) «Основы BigData и анализа данных»** является изучение основных методов анализа данных с использованием современных компьютерных технологий в задачах поиска информации, обработки и анализа данных, а также приобретение навыков исследователя данных (datascientist).

### **1.2. Задачи освоения дисциплины (модуля):**

- сформировать у студентов представления о массивах больших данных и основных методах анализа данных;
- научить основным приемам получения данных;
- научить адекватному выбору метода анализа при решении конкретных задач;
- сформировать умения и навыки использования цифровых технологий в профессиональной деятельности.

## **2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП**

**2.1. Учебная дисциплина (модуль) «Основы BigData и анализа данных»** относится к части, формируемой участниками образовательных отношений и осваивается в 3 семестре.

**2.2. Для изучения данной учебной дисциплины (модуля) необходимы следующие знания, умения, навыки, формируемые предшествующими дисциплинами:**

- Высшая математика, Информатика.

#### **Знания:**

- различные методы и методики получения и проведения анализа метаданных.

#### **Умения:**

- интерпретировать полученные в ходе анализа результаты с точки зрения подготовки управленческих решений.

#### **Навыки:**

- владеть основами работы с большими массивами данных (BigData).

**2.3. Последующие учебные дисциплины (модули) и (или) практики, для которых необходимы знания, умения, навыки, формируемые данной учебной дисциплиной (модулем):**

- специальные и профессиональные дисциплины,
- выпускная квалификационная работа.

## **3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

Процесс освоения дисциплины (модуля) направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки / специальности:

- а) универсальной (УК): УК-1.

**Таблица 1 – Декомпозиция результатов обучения**

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю)		
	Знать (1)	Уметь (2)	Владеть (3)
<p>УК-1</p> <p>Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач</p>	<p>ИУК 1.1.1</p> <p>Анализирует задачу, выделяя ее базовые составляющие, осуществляет декомпозицию задачи</p>	<p>ИУК 1.2.2</p> <p>Находит и критически анализирует информацию, необходимую для решения поставленной задачи</p> <p>ИУК 1.2.3</p> <p>Рассматривает возможные варианты решения задачи, оценивая их достоинства и недостатки</p>	<p>ИУК 1.3.4</p> <p>Грамотно, логично, аргументированно формирует собственные суждения и оценки. Отличает факты от мнений, интерпретаций, оценок и т.д. в рассуждениях других участников деятельности</p> <p>ИУК 1.3.5</p> <p>Определяет и оценивает последствия возможных решений задачи</p>

#### 4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объем дисциплины (модуля) составляет 2 зачетные единицы, в том числе 36 часов, выделенных на контактную работу обучающихся с преподавателем (из них 36 часов – практические, семинарские занятия), и 36 часов – на самостоятельную работу обучающихся.

**Таблица 2 - Структура и содержание дисциплины (модуля)**

№ п/п	Наименование раздела (темы)	Семестр	Контактная работа (в часах)			Самостоят. работа		Формы текущего контроля успеваемости, форма промежуточной аттестации [по семестрам]
			Л	ПЗ	ЛР	КР	СР	
1	Введение в BigData и анализ данных: основные понятия, области применения, современных технологий обработки и анализа больших данных.	3		4			4	Устный опрос.
2	Большие данные. Организация сбора и хранения больших наборов данных.			4			4	Устный опрос. Практическое задание.
3	Основы математической статистики. STATISTICA. Настройка Excel «Пакет анализа».			10			10	Устный опрос. Практическое задание.
4	Предварительный и разведочный анализ данных в Рили Python.			10			10	Устный опрос. Практическое задание.

5	Визуализация данных и результатов анализа. Выводы и рекомендации по организации дальнейшей работы.			8			8	Устный опрос. Практическое задание.
<b>ИТОГО</b>				<b>36</b>			<b>36</b>	<b>ЗАЧЕТ</b>

**Таблица 3 – Матрица соотнесения разделов, тем учебной дисциплины (модуля) и формируемых компетенций**

Раздел, тема дисциплины (модуля)	Кол-во часов	Код компетенции	Общее количество компетенций
		УК-1	
Введение в BigData и анализ данных: основные понятия, области применения, современных технологий обработки и анализа больших данных.	8	+	1
Большие данные. Организация сбора и хранения больших наборов данных.	8	+	1
Основы математической статистики. STATISTICA. Настройка Excel «Пакет анализа».	20	+	1
Предварительный и разведочный анализ данных в R или Python.	20	+	1
Визуализация данных и результатов анализа. Выводы и рекомендации по организации дальнейшей работы.	16	+	1
<b>ИТОГО</b>	<b>72</b>		<b>1</b>

#### **Краткое содержание каждой темы дисциплины (модуля)**

##### **Тема 1. Введение в BigData и анализ данных: основные понятия, области применения современных технологий обработки и анализа больших данных**

Предмет и задачи курса. Терминология, история появления. Технические сложности работы с большими данными (BigData). Большие данные в информационных технологиях: совокупность подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов. Данные и метаданные.

##### **Тема 2. Большие данные. Организация сбора и хранения больших наборов данных.**

Сбор, хранение, обработка, извлечение и анализ больших потоков данных. Основные характеристики и типы больших данных. Формы представления данных, типы и виды данных. Представления наборов данных.

##### **Тема 3. Основы математической статистики. STATISTICA. Настройка Excel «Пакет анализа».**

Основные аналитические методы обработки данных. Процесс анализа. Общая схема анализа. Программное обеспечение в области анализа данных. Статистический подход к анализу данных. Точечное и интервальное оценивание. Выявление взаимосвязей. Корреляционный анализ. Регрессия.

#### **Тема 4. Предварительный (разведочный) анализ данных в R или Python.**

Подготовка данных к анализу. Работа с выбросами и пропущенными значениями. Предварительное исследование Dataset: определение его основных характеристик, взаимосвязей между признаками. Обработка пропусков.

#### **Тема 5. Визуализация данных и результатов анализа.**

Важность визуализации данных. Основные принципы визуализации. Способы визуализации данных. Презентация результатов.

### **5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРЕПОДАВАНИЮ И ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)**

#### **5.1. Указания для преподавателей по организации и проведению учебных занятий по дисциплине (модулю)**

При подготовке к занятию преподаватель должен обратить особое внимание на целевые установки студентам, учебные и воспитательные цели занятия. Преподаватель должен обязательно ознакомиться с рекомендованной для обучаемых литературой, тщательно изучить применяемые наглядные пособия.

План проведения занятий должен включать: учебные вопросы при традиционных методах обучения; вопросы проблемного характера и проблемные ситуации, если применяется проблемный метод изложения; путеводитель предъявления иллюстративного материала; цитаты, примеры, определения, формулировки; другие элементы по желанию преподавателя.

В процессе обучения студентов предусматривается применение интерактивных и иных форм инновационных учебных занятий, развивающих у обучающихся навыки командной работы, межличностной коммуникации, принятия решений, лидерские качества (включая проведение интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализ ситуаций и имитационных моделей, преподавание дисциплин в форме курсов, составленных на основе результатов научных исследований, проводимых университетом, в том числе с учетом потребностей работодателей). Реализация компетентного подхода предусматривает широкое использование в учебном процессе активных и интерактивных форм проведения занятий (компьютерных симуляций, деловых и ролевых игр, разбор конкретных ситуаций, дискуссии) в сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся.

#### **5.2. Указания для обучающихся по освоению дисциплины (модулю)**

Основу теоретического обучения студентов составляют систематизированные знания студентам о наиболее сложных и актуальных проблемах изучаемой дисциплины. На занятиях особое внимание уделяется не только усвоению студентами изучаемых проблем, но и стимулированию их активной познавательной деятельности, творческого мышления, развитию научного профессионально-значимых свойств и качеств.

Осуществляя учебные действия на занятиях, студенты должны внимательно воспринимать действия преподавателя, запоминать складывающиеся образы, мыслить, добиваться понимания изучаемого предмета, применения знаний на практике, при решении учебно-профессиональных задач. Студенты должны аккуратно вести конспект. В случае недопонимания какой-либо части предмета следует задать вопрос в установленном порядке преподавателю. В процессе работы необходимо так же выполнять в конспектах модели изучаемого предмета (рисунки, схемы, чертежи и т.д.), которые использует преподаватель.

Работу над конспектом следует начинать с его доработки, желательно в тот же день, пока материал еще легко воспроизводим в памяти (через 10 часов после лекции в

памяти остается не более 30-40 % материала). Записи должны быть наглядными, для чего следует применять различные способы выделений. В ходе доработки конспекта углубляются, расширяются и закрепляются знания, а также дополняется, исправляется и совершенствуется конспект.

Непременным условием глубокого усвоения учебного материала является знание основ, на которых строится изложение материала. Обычно преподаватель напоминает, какой ранее изученный материал и в какой степени требуется подготовить к очередному занятию. Обращение к ранее изученному материалу не только помогает восстановить в памяти известные положения, выводы, но и приводит разрозненные знания в систему, углубляет и расширяет их.

При работе с основной и дополнительной литературой придерживаться такой последовательности. Сначала прочитать весь заданный текст в быстром темпе. Цель такого чтения заключается в том, чтобы создать общее представление об изучаемом материале, понять общий смысл прочитанного. Затем прочитать вторично, более медленно, чтобы в ходе чтения понять и запомнить смысл каждой фразы, каждого положения и вопроса в целом.

Подготовка докладов, выступлений и рефератов. Реферат представляет письменный материал по определённой теме, в котором собрана информация из одного или нескольких источников. В нем в обобщенном виде представляется материал на определенную тему, включающий обзор соответствующих литературных и других источников. Рефераты могут являться изложением содержания какой-либо научной работы, статьи и т.п. Доклад представляет публичное, развернутое сообщение (информирование) по определенному вопросу или комплексу вопросов, основанное на привлечении документальных данных, результатов исследования, анализа деятельности и т.д.

При подготовке к докладу на семинаре по теме, указанной преподавателем, студент должен ознакомиться не только с основной, но и дополнительной литературой, а также с последними публикациями по этой тематике в сети Интернет. Необходимо подготовить текст доклада и иллюстративный материал в виде презентации. Доклад должен включать введение, основную часть и заключение. На доклад отводится 20-25 минут учебного времени. Он должен быть научным, конкретным, определенным, глубоко раскрывать проблему и пути ее решения. Особенно следует обратить внимание на безусловную обязательность решения домашних задач, указанных преподавателем к семинару.

**Таблица 4 – Содержание самостоятельной работы обучающихся**

<i>Номер раздела (темы)</i>	<i>Темы/вопросы, выносимые на самостоятельное изучение</i>	<i>Кол-во часов</i>	<i>Формы работы</i>
1	Поиск данных по заданию преподавателя	4	Проектное задание.
2	Сбор и представление найденных данных в виде Dataframe	4	Проектное задание.
3	Разведочный анализ полученных данных	10	Проектное задание.
4	Установление взаимосвязей и выводы	10	Проектное задание.
5	Визуализация результатов	8	Проектное задание.

### **5.3. Виды и формы письменных работ, предусмотренных при освоении дисциплины, выполняемые обучающимися самостоятельно.**

#### **Критерии оценивания компетенций (результатов)**

- правильное изложения доклада (актуальность, цель и задачи, содержание, выводы, рекомендации) и составление презентации (оформление слайдов, информационное содержание, список источников научной литературы);
- уровень раскрытия темы;
- тематическое соответствие содержания доклада и презентации;
- количество использованных источников научной литературы;
- уровень владения материалом и терминологией по дисциплине в ходе защиты курсовой работы.

## 6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

### 6.1. Образовательные технологии

Структура освоения дисциплины «Основы BigData и анализ данных» предусматривает использование следующих образовательных технологий по видам учебных работ:

Практические занятия с использованием мультимедийного оборудования.

Включение бесплатного онлайн курса «BigData и Data Science: начни погружение с нуля» на образовательной платформе: Stepik- <https://stepik.org/course/101687/promo>

Дистанционные формы выполнения заданий.

**Практические занятия.** Основной формой является выполнение практических работ, знакомство со специализированными приемами и программами работы с данными в интернете, в программах MS Excel и STATISTICA и языками программирования R и Python.

Активные и интерактивные формы обучения включают: собеседование по поиску и сбору требуемых данных, алгоритму обработки, анализу, обсуждения результатов обработки информации (работа в команде и метод проектов). Анализ, обобщение материалов по заданиям, а также просмотр и обобщение материалов презентаций.

На занятиях используются:

- 1) учебно-наглядные материалы: фото-слайды, формулы, выборки;
- 2) специализированные программные продукты.

**Компьютерные симуляции** в форме решения задач оценки соответствия полученных результатов гипотезе нормального распределения, характеристики вариабельности выборок, проверки гипотез распределения и оценки связи между признаками.

**Дискуссии** о достоинствах и недостатках применения параметрических и непараметрических способов обработки полученных результатов, выборах основных факторов, влияющих на распределение полученных результатов при использовании дисперсионного анализа обработки данных.

**Таблица 5 – Образовательные технологии, используемые при реализации учебных занятий**

Раздел, тема дисциплины (модуля)	Форма учебного занятия		
	Лекция	Практическое занятие, семинар	Лабораторная работа
Введение в BigData и анализ данных: основные понятия, области применения, современных технологий обработки и анализа больших данных.	<i>Не предусмотрено</i>	<i>Выполнение практических заданий. Мини-проект индивидуальный. Групповая работа</i>	<i>Не предусмотрено</i>
Большие данные. Организация сбора и хранения больших наборов данных.	<i>Не предусмотрено</i>	<i>Мини-проект индивидуальный.</i>	<i>Не предусмотрено</i>
Основы математической ста-	<i>Не предусмотрено</i>	<i>Устный опрос. Ми-</i>	<i>Не предусмотрено</i>

тистики. STATISTICA. Надстройка Excel «Пакет анализа».	<i>но</i>	<i>ни-проект. Групповая работа</i>	<i>рено</i>
Предварительный и разведочный анализ данных в R или Python.	<i>Не предусмотрено</i>	<i>Устный опрос. Мини-проект индивидуальный. Групповая работа</i>	<i>Не предусмотрено</i>
Визуализация данных и результатов анализа. Выводы и рекомендации по организации дальнейшей работы.	<i>Не предусмотрено</i>	<i>Устный опрос. Мини-проект индивидуальный. Групповая работа</i>	<i>Не предусмотрено</i>

## 6.2. Информационные технологии

— Использование информационного ресурса преподавателя (рассылка заданий, предоставление выполненных работ, ответы на вопросы, ознакомление учащихся с оценками);

— Включение бесплатного онлайн курса «BigData и DataScience: начни погружение с нуля» на образовательной платформе: Stepik - <https://stepik.org/course/101687/promo>;

— -Открытые видеолекции учебных курсов МГУ <https://teach-in.ru/course/probability-theory-and-mathematical-statistics-for-geologists>

— Использование тематических сайтов по статистике и анализу данных, языку программирования R: <https://r-analytics.blogspot.com>, <http://window.edu.ru/catalog/resources/uchebnik-statistika>, <https://rstudio-pubs-static.s3.amazonaws.com>,

— Использование образовательного портала АГУ <http://learn.asu.edu.ru/>

— Электронная библиотека «Астраханский государственный университет» собственной генерации на платформе ЭБС «Электронный Читальный зал – BiblioТех». <https://biblio.asu.edu.ru>. Учетная запись образовательного портала АГУ

— Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». Многопрофильный образовательный ресурс «Консультант студента» является электронной библиотечной системой, предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Каталог содержит более 15 000 наименований изданий. [www.studentlibrary.ru](http://www.studentlibrary.ru). Регистрация с компьютеров АГУ

— Электронная библиотечная система издательства ЮРАЙТ, раздел «Легендарные книги». [www.biblio-online.ru](http://www.biblio-online.ru), <https://urait.ru/>

— Электронная библиотечная система IPRbooks. [www.iprbookshop.ru](http://www.iprbookshop.ru)

— Электронно-библиотечная система BOOK.ru

— Использование виртуальной обучающей среды (LMS Moodle «Электронное образование»)

— Использование средств представления учебной информации для проведения лекций и семинаров с использованием презентаций

## 6.3. Программное обеспечение, современные профессиональные базы данных и информационные справочные системы

### 6.3.1. Программное обеспечение

Наименование программного обеспечения	Назначение
Adobe Reader	Программа для просмотра электронных документов
Платформа дистанционного обучения LMS Moodle	Виртуальная обучающая среда

Наименование программного обеспечения	Назначение
Mozilla Fire Fox	Браузер
Microsoft Office 2013, Microsoft Office Project 2013, Microsoft Office Visio 2013	Пакет офисных программ
7-zip	Архиватор
Microsoft Windows 7 Professional	Операционная система
Kaspersky Endpoint Security	Средство антивирусной защиты
Google Chrome	Браузер
Notepad++	Текстовый редактор
Open Office	Пакет офисных программ
Opera	Браузер
Paint .NET	Растровый графический редактор
Microsoft Security Assessment Tool. Режим доступа: <a href="http://www.microsoft.com/ru-ru/download/details.aspx?id=12273">http://www.microsoft.com/ru-ru/download/details.aspx?id=12273</a> (Free) Windows Security Risk Management Guide Tools and Templates. Режим доступа: <a href="http://www.microsoft.com/en-us/download/details.aspx?id=6232">http://www.microsoft.com/en-us/download/details.aspx?id=6232</a> (Free)	Программы для информационной безопасности
VLC Player	Медиапроигрыватель
Win DjView	Программа для просмотра файлов в формате DJV и DjVu

### 6.3.2. Современные профессиональные базы данных и информационные справочные системы

- Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем». <https://library.asu.edu.ru>
- Электронный каталог «Научные журналы АГУ»: <http://journal.asu.edu.ru/>
- Универсальная справочно-информационная полнотекстовая база данных периодических изданий ООО "ИВИС". <http://dlib.eastview.com>  
Имя пользователя: AstrGU  
Пароль: AstrGU
- Электронно-библиотечная система eLibrary. <http://elibrary.ru>
- Корпоративный проект Ассоциации региональных библиотечных консорциумов (АРБИКОН) «Межрегиональная аналитическая роспись статей» (МАРС) - сводная база данных, содержащая полную аналитическую роспись 1800 названий журналов по разным отраслям знаний. Участники проекта предоставляют друг другу электронные копии отсканированных статей из книг, сборников, журналов, содержащихся в фондах их библиотек. <http://mars.arbicon.ru>
- Справочная правовая система КонсультантПлюс. Содержится огромный массив справочной правовой информации, российское и региональное законодательство, судебную практику, финансовые и кадровые консультации, консультации для бюджетных организаций, комментарии законодательства, формы документов, проекты нормативных правовых актов, международные правовые акты, правовые акты, технические нормы и правила. <http://www.consultant.ru>
- Информационно-правовое обеспечение «Система ГАРАНТ». В системе ГАРАНТ представлены федеральные и региональные правовые акты, судебная практика, книги, энциклопедии, интерактивные схемы, комментарии ведущих специалистов и мате-

риалы известных профессиональных изданий, бланки отчетности и образцы договоров, международные соглашения, проекты законов.

- Предоставляет доступ к федеральному и региональному законодательству, комментариям и разъяснениям из ведущих профессиональных СМИ, книгам и обновляемым энциклопедиям, типовым формам документов, судебной практике, международным договорам и другой нормативной информации. Всего в нее включено более 2,5 млн документов. В программе представлены документы более 13 000 федеральных, региональных и местных эмитентов. <http://garant-astrakhan.ru>
- Единое окно доступа к образовательным ресурсам <http://window.edu.ru>
- Министерство науки и высшего образования Российской Федерации <https://minobrnauki.gov.ru/>
- Министерство просвещения Российской Федерации <https://edu.gov.ru>
- Официальный информационный портал ЕГЭ <http://www.ege.edu.ru>
- Федеральное агентство по делам молодежи (Росмолодежь) <https://fadm.gov.ru>
- Федеральная служба по надзору в сфере образования и науки (Рособрнадзор) <http://obrnadzor.gov.ru>
- Сайт государственной программы Российской Федерации «Доступная среда» <http://zhit-vmeste.ru>
- Российское движение школьников <https://рдш.рф>
- Официальный сайт сетевой академии cisco: [www.netacad.com](http://www.netacad.com)

## 7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

### 7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине (модулю) «Основы BigData и анализа данных» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины (модуля) – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

**Таблица 6 - Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств**

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Введение в BigData и анализ данных: основные понятия, области применения, современных технологий обработки и анализа больших данных.	УК-1	Устный опрос.
2	Большие данные. Организация сбора и хранения больших наборов данных.	УК-1	Устный опрос. Практическое задание.
3	Основы математической статистики. STATISTICA. Настройка Excel «Пакет анализа».	УК-1	Устный опрос. Практическое задание.
4	Предварительный и разведочный анализ данных в Рили Python.	УК-1	Устный опрос. Практическое задание.
5	Визуализация данных и результатов анализа. Выводы и рекомендации по организации	УК-1	Устный опрос. Практическое задание.

	дальнейшей работы.		
--	--------------------	--	--

Для оценивания результатов обучения в виде знаний используются следующие типы контроля:

- устный опрос;
- письменные ответы на вопросы;
- проектные задания.

Данные типы контроля проводятся по разработанным вопросам по отдельному учебному элементу программы.

Для оценивания результатов обучения в виде умений и владений используются следующие типы контроля:

- практические задания, включающие одну или несколько задач в виде краткой формулировки действий (комплекса действий), которые следует выполнить и описать полученный результат (сделать вывод).

Оценка качества подготовки обучающихся включает текущую и промежуточную аттестацию знаний – зачет и экзамен. Данные типы контроля служат основным средством обеспечения в учебном процессе «обратной связи» между преподавателем и студентом, необходимой для стимулирования работы обучающихся и совершенствования методики преподавания учебной дисциплины.

Зачет по дисциплине «Основы BigData и анализ данных» сдаётся в письменной форме по билетам. При этом студент должен ответить на 1 вопрос из примерного перечня вопросов и выполнить 1 практическую задачу.

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

**Таблица 7 - Показатели оценивания результатов обучения в виде знаний**

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении, затрудняется в приведении примеров и формулировке выводов
2 «неудовлетворительно»	демонстрирует существенные пробелы в знании теоретического материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры

**Таблица 8 - Показатели оценивания результатов обучения в виде умений и владений**

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4	демонстрирует способность применять знание теоретического материала

«хорошо»	при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, не способен применить знание теоретического материала при выполнении заданий, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание при подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задание

### **7.3. Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности**

#### **Тема 1. Введение в BigData и анализ данных: основные понятия, области применения современных технологий обработки и анализа больших данных.**

##### **Вопросы для обсуждения:**

1. Что такое Big Data и Data Science
2. Чем занимаются специалисты по анализу данных и почему они так востребованы сейчас.
3. Признаки и их классификация данных.
4. Данные и метаданные.
5. Что такое совокупность?
6. Области применения анализа метаданных

#### **Тема 2. Большие данные. Организация сбора и хранения больших наборов данных.**

##### **Вопросы для обсуждения:**

1. Источники больших данных
2. Как искать большие данные?
3. Организация сбора больших данных
4. Хранение больших данных

##### **Практическая работа № 1: «Поиск больших данных»:**

1. Найти в сети интернет или других источниках и составить массив данных одного типа объемом 100 единиц, с которыми можете столкнуться в своей будущей профессии.
2. Сохранить полученный результат для дальнейшей работы в программном пакете Microsoft Excel.

#### **Тема 3. Основы математической статистики. STATISTICA. Настройка Excel «Пакет анализа».**

##### **Вопросы для обсуждения:**

1. Что такое анализ больших данных
2. Пакеты для обработки данных
3. Научные проблемы и большие данные
4. Использование специализированных программных пакетов для анализа данных

**Практическая работа № 2: «Расчет статистических параметров»:**

1. Сгруппировать полученные данные в вариационный ряд и рассчитать основные статистические параметры в программном пакете Microsoft Excel.
2. Сделайте выводы о характере распределения данных внутри массива.

**Тема 4. Предварительный (разведочный) анализ данных в R или Python.**

**Практическая работа № 3 «Обработка данных»:**

1. Для полученного массива данных провести разведочный анализ данных и проанализировать данные с целью выявления взаимосвязи, взаимозависимости и зависимости от внешних факторов.
2. Провести регрессионный анализ с целью прогнозирования дальнейшего поведения данных. Сделайте выводы.

**Тема 5. Визуализация данных и результатов анализа. Выводы и рекомендации по организации дальнейшей работы.**

**Вопросы для обсуждения:**

1. Типы представления данных.
2. Выбор оптимального типа визуализации
3. Презентация результатов

**Практическая работа № 4: «Визуализация данных»:**

1. Визуализировать результаты практической работы № 2 и 3 с использованием программных пакетов Microsoft Excel и STATISTICA, язык программирования R.
2. Построить диаграммы размаха (Box&Wiskers Plot), гистограммы, нормальные вероятностные графики (Normal Probability plot). Описать и сделать выводы.

**Вопросы для подготовки к зачету по дисциплине**

1. Понятия об испытаниях, событиях и величинах.
2. Признаки и их классификация.
3. Генеральная совокупность и выборка.
4. Статистический программный пакет STATISTICA v.6.
5. Программный пакет MS Excel.
6. Теория вероятности.
7. Закон нормального распределения.
8. Нормированного отклонения.
9. Основные статистические параметры.
10. Роль среднего арифметического.
11. Роль и статистический смысл дисперсии и коэффициента вариации в статистической обработке данных.
12. Оценки моды, медианы и среднего арифметического.
13. Оценка дисперсии, стандартного отклонения и коэффициентов вариации.
14. Ошибки среднего, дисперсии, стандартного отклонения и коэффициента вариации.
15. Статистические гипотезы.
16. Параметрические методы проверки статистических гипотез.
17. Сравнение двух и нескольких дисперсий. Сравнение двух средних.
18. Критерий Стьюдента. Критерия Фишера. t-тест.
19. Корреляция. Коэффициент корреляции. Корреляционный анализ.

20. Регрессия. Коэффициент регрессии. Регрессионный анализ.
21. Дисперсионный анализ. Одно- и двухфакторный дисперсионный анализ.
22. Непараметрические методы математической статистики.
23. Критерии Вилкоксона, Манн-Уитни, Колмогорова-Смирнова.
24. Понятие Большие данные (BigData).
25. Роль цифровой информации в 21 веке.
26. Проблемы анализа и обработки большого объема данных.
27. Базовые принципы обработки больших данных.
28. Подготовка данных к анализу.
29. Общая схема анализа данных.
30. Получение Dataset и преобразование в Dataframe.
31. Получение информации по метаданным.
32. Установление взаимозависимостей.
33. Принятие управленческих решений на основе анализа данных
34. Типы визуализации данных.
35. Представление результатов анализа данных.
36. Роль BigData и анализа данных в различных сферах жизнедеятельности человека.

ка.

**Таблица 9 – Примеры оценочных средств с ключами правильных ответов**

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
<b><i>УК-1 способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач</i></b>				
1.	Задание закрытого типа	Если значение коэффициента корреляции составляет 0,8 то связь между взаимосвязанными признаками <ol style="list-style-type: none"> <li>1. умеренная</li> <li>2. заметная</li> <li>3. слабая</li> <li>4. тесная</li> </ol>	4	1
2.		Серединное по номеру значение ранжированного ряда данных, которое разбивает ряд на по количеству вариант части: <ol style="list-style-type: none"> <li>1. Медиана</li> <li>2. Мода</li> <li>3. Среднее арифметическое</li> <li>4. Среднее взвешенное</li> </ol>	1	1
3.		Укажите источники больших данных: <ol style="list-style-type: none"> <li>1. Статистика поведения пользователей на сайте и в приложении;</li> <li>2. Социальные опросы;</li> <li>3. Записи с камер ви-</li> </ol>	1, 2, 3	2

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		деонаблюдения		
4.		Визуализация аналитических данных — это	Анимированные модели или графики, созданные на основе больших данных.	3
5.		В каких отраслях уже используют Big Data? 1. Промышленность 2. Медицина 3. Интернет вещей	1, 2, 3	2
6.	Задание открытого типа	Посредством каких показателей анализируют данные?	Данные анализирую преимущественно с помощью статистических показателей. Эти показатели должны быть рассчитаны через количественно – качественные характеристики какого – то свойства группы единиц совокупности или совокупности в целом.	3
7.		В чем суть Достоверности данных?	Достоверность подразумевает полноту и общую точность данных. Именно на основе достоверных данных организации и компании принимают важные решения. Тем не менее стопроцентная достоверность требуется не всегда. Так, это некритично для часто обновляемых данных из соцсетей. Даже если несколько десятков или сотен записей окажутся неверными, на общую статистику это не повлияет.	5
8.		Каков критерий объема БОЛЬШИХ данных?	Термин «большие данные» предложил редактор журнала Nature Клиффорд Линч в спецвыпуске 2008 года. Он говорил о взрывном росте объемов информации в мире. К большим данным Линч отнес любые массивы неоднородных данных более 150 Гб в сутки, однако единого критерия до сих пор не существует.	5
9.		Для чего структурируют большие данные?	BigData структурируют сначала для того, чтобы отобрать только те, что нужны для анализа.	3
10.		Является ли необходимой при анализе данных Описательная аналитика (descriptive analytics) и почему?	Описательная аналитика (descriptive analytics) — самый распространенный элемент анализа данных. Она отвечает на вопрос «Что произошло?», анализирует данные, поступа-	5

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
			ющие в реальном времени, и исторические данные. Главная цель — выяснить причины и закономерности успехов или неудач в той или иной сфере, чтобы использовать эти данные для наиболее эффективных моделей. Для описательной аналитики используют базовые математические функции. Типичный пример — социологические исследования или данные веб-статистики, которые компания получает через Google Analytics.	

Полный комплект оценочных материалов по дисциплине (модулю) (фонд оценочных средств) хранится в электронном виде на кафедре, утверждающей рабочую программу дисциплины (модуля), и в Центре мониторинга и аудита качества обучения.

#### 7.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Оценка результатов и обучения студента выполняется в соответствии с «Положением об балльно-рейтинговой системе оценки учебных достижений студентов», утвержденным решением Ученого совета ФГБОУ «Астраханский государственный университет» от 30 декабря 2013г.

**Таблица 10 – Технологическая карта рейтинговых баллов по дисциплине (модулю)**

№ п/п	Контролируемые мероприятия	Количество мероприятий/ баллы	Максимальное количество баллов	Срок предоставления
<b>Основной блок</b>				
1	Полный ответ по вопросу	9 (36)	27	по расписанию
2	Дополнение	9 (16)	9	по расписанию
3	Выполнение практического задания	9 (56)	45	по расписанию
4	Выполнение индивидуального задания	9(16)	9	Указан в Moodle
<b>Всего</b>			<b>90</b>	
<b>Блок бонусов</b>				
3	Посещение занятий	9 (0,56)	5	по расписанию
4	Активность студента на занятии	9 (0,56)	5	
<b>Всего</b>			<b>10</b>	
<b>Дополнительный блок</b>				
5	Зачет			по расписанию
<b>Всего</b>				
<b>ИТОГО</b>			<b>100</b>	

**Таблица 11 – Система штрафов (для одного занятия)**

Показатель	Балл
Опоздание (два и более)	-2
Не готов к практической части занятия	-3
Нарушение учебной дисциплины	-2
Пропуск занятия без уважительной причины (за одно занятие)	-4
Нарушение правил техники безопасности	-1

**Таблица 12 – Шкала перевода рейтинговых баллов в итоговую оценку за семестр по дисциплине (модулю)**

Сумма баллов	Оценка по 4-балльной шкале	
90–100	5 (отлично)	Зачтено
85–89	4 (хорошо)	
75–84		
70–74		
65–69	3 (удовлетворительно)	
60–64		
Ниже 60	2 (неудовлетворительно)	Не зачтено

При реализации дисциплины (модуля) в зависимости от уровня подготовленности обучающихся могут быть использованы иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

## **8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)**

### **8.1. Основная литература**

1. Колесниченко О.Ю. Data Science (наука о данных) в становлении информационного общества : учебное пособие / Колесниченко О.Ю.. — Москва : Прометей, 2021. — 52 с. — ISBN 978-5-00172-110-9. — Текст : электронный // IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/125600.html>
2. Пашкевич, О. И. Статистическая обработка эмпирических данных в системе STATISTICA : учебно-методическое пособие / О.И. Пашкевич. — Минск : Республиканский институт профессионального образования (РИПО), 2014. — 148 с. — ISBN 978-985-503-385-2. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/67607.html>
3. Практическая статистика для специалистов DataScience: Пер. с англ. /П. Брюс, Э. Брюс, П. Гедек. — 2-е изд., перераб. и доп. — СПб.:БХВ-Петербург, 2021. — 352 с.: ил. (Адаптированный вариант примеров в виде электронного архива - <ftp://ftp.bhv.ru/9785977567053.zip>).
4. Руководство по изучению языка R и его использование в Data Science: <https://proglib.io/p/data-science-with-r>
5. Учебно-методическое пособие по программе STATISTICA – <http://statosphere.ru/books-arch/statistica-books/88-uch-met-pos.html>
6. Халеева Е.П. Анализ данных средствами языка R : учебное пособие / Халеева Е.П., Аль-Ханани М.А., Лютикова М.Н.. — Саратов: Вузовское образование, 2022. — 71 с. — ISBN 978-5-4487-0824-4. — Текст : электронный // IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/119442.html>

## 8.2. Дополнительная литература

1. Васильева, Э. К. Статистика : учебник для студентов вузов, обучающихся по специальностям экономики и управления (080100) / Э. К. Васильева, В. С. Лялин. — Москва : ЮНИТИ-ДАНА, 2012. — 398 с. — ISBN 978-5-238-01192-9. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/8581.html>
2. Зуев С.В. Методы анализа данных : учебное пособие / Зуев С.В.. — Москва : Ай Пи Ар Медиа, 2023. — 132 с. — ISBN 978-5-4497-2107-5. — Текст : электронный // IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/129059.html>
3. Илышев, А. М. Общая теория статистики : учебник / А. М. Илышев. — Москва : ЮНИТИ-ДАНА, 2012. — 536 с. — ISBN 978-5-238-01446-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/10504.html>
4. Медведев П.В., Математическая обработка результатов исследования : учебное пособие / Медведев П.В. - Оренбург: ОГУ, 2017. - 99 с. - ISBN 978-5-7410-1772-2 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785741017722.html>
5. Новикова, Е. Н. Компьютерная обработка результатов измерений : учебное пособие / Е. Н. Новикова, О. Л. Серветник. — Ставрополь : Северо-Кавказский федеральный университет, 2017. — 182 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/75577.html>
6. Понкратова, Т. А. Статистика. Общая теория статистики : учебное пособие / Т. А. Понкратова, О. С. Кузнецова, О. В. Секлецова. — Кемерово : Кемеровский технологический институт пищевой промышленности, 2011. — 143 с. — ISBN 978-5-89289-668-9. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/14390.html>
7. Шпаков, П. С. Математическая обработка результатов измерений : учебное пособие / П. С. Шпаков, Ю. Л. Юнаков. — Красноярск : Сибирский федеральный университет, 2014. — 410 с. — ISBN 978-5-7638-3077-4. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/84372.html>

## 8.3 Интернет-ресурсы, необходимые для освоения дисциплины (модуля)

1. Электронная библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента»: [www.studentlibrary.ru](http://www.studentlibrary.ru).
2. Электронно-библиотечная система IPR BOOKS: <http://www.iprbookshop.ru>
3. Электронная библиотека АГУ <https://biblio.asu.edu.ru>

## 9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для проведения практических занятий необходим один дисплейный класс ПК, объединенных в локальную вычислительную сеть с доступом в интернет и оборудованный мультимедийной установкой (проектором). Для проведения лекционных занятий – аудитория, оборудованная мультимедийной установкой.

Для проведения занятий по дисциплине имеются аудитории для проведения лекционных занятий, оборудованные мультимедийной техникой с возможностью презентации обучающих материалов, фрагментов фильмов; аудитории для проведения семинарских и практических занятий, оборудованные учебной мебелью и средствами наглядного представления учебных материалов; библиотека с местами, оборудованными компьютерами, имеющими доступ к сети Интернет.

Рабочая программа дисциплины (модуля) при необходимости может быть адаптирована для обучения (в том числе с применением дистанционных образовательных техно-

логий) лиц с ограниченными возможностями здоровья, инвалидов. Для этого требуется заявление обучающихся, являющихся лицами с ограниченными возможностями здоровья, инвалидами, или их законных представителей и рекомендации психолого-медико-педагогической комиссии. Для инвалидов содержание рабочей программы дисциплины (модуля) может определяться также в соответствии с индивидуальной программой реабилитации инвалида (при наличии).