

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ
Заведующий кафедрой
Регионоведения и экономики
зарубежных стран



Беленов О.Н.
20.03.2024

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.ДВ.02.01 «Анализ больших данных»

Код и наименование дисциплины в соответствии с учебным планом

- 1. Шифр и наименование направления подготовки/специальности:**
41.03.01 Зарубежное регионоведение
- 2. Профиль подготовки/специализации:** Евразийские исследования
- 3. Квалификация (степень) выпускника:** бакалавр
- 4. Форма образования:** очная
- 5. Кафедра, отвечающая за реализацию дисциплины:** регионоведения и экономики зарубежных стран
- 6. Составители программы:** Шилова Ирина Валерьевна, преподаватель
- 7. Рекомендована:** НМС факультета МО протокол НМС 20.03.2024 №3
- 8. Учебный год:** 2028-2029 **Семестр:** 8

9. Цели и задачи учебной дисциплины: формирование у студентов профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных. Приобретенные знания позволят сформировать у студентов практические навыки по: постановке задач анализа данных; предварительной обработке данных; визуализации данных; применению методов интеллектуального анализа данных к большим массивам данных.

Задачи учебной дисциплины:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- сформировать представления о базовых принципах сбора, обработки и анализа маркетинговой информации для принятия управленческих решений.

10. Место учебной дисциплины в структуре ООП: Дисциплина относится к блоку Б1 учебного плана, включена в его вариативную часть, является курсом по выбору. Обучающийся должен обладать знаниями по экономической теории, офисным информационным технологиям. В части программного обеспечения необходимо владение MS Excel.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников):

Код	Название компетенции	Код	Индикатор(ы)	Планируемые результаты обучения
(ПК-1)	Способен формировать возможные решения на основе разработанных для них целевых показателей	ПК - 1.1	Способен осуществлять выявление, сбор и анализ информации бизнес-анализа для формирования возможных решений	знать: стандарты, модели и алгоритмы обработки и анализа больших данных уметь: формально описывать задачи, возникающие в бизнес-аналитике, и сводить их к математическим или технологическим задачам владеть: навыками обработки экономической информации с применением IT-технологий

12. Объем дисциплины в зачетных единицах/час. (в соответствии с учебным планом) — 3/108.

Форма промежуточной аттестации (зачет/экзамен) зачет.

13. Виды учебной работы

Вид учебной работы	Трудоемкость	
	Всего	По семестрам
		5
Аудиторные занятия	50	50
в том числе: лекции	16	16
практические	34	34
лабораторные		
Самостоятельная работа	58	58
Форма промежуточной аттестации (зачет – 0 час. / экзамен – __ час.)		
Итого:	108	108

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины
Лабораторные занятия		
1	Введение в анализ больших данных. Обзор источников информации.	Основные определения, термины, задачи анализа больших данных. Понятие Data Mining. Когнитивный анализ данных. Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований). Методики сбора данных.
2	Технологии хранения и обработки больших данных.	Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных
3	Статистические методы анализа данных.	Основные понятия математической статистики. Методы анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы
4	Современные программные средства анализа больших данных.	Обзор современных популярных программных средства анализа данных: Statistica, SPSS, Excel, R-Studio и другие; их преимущества и недостатки.
5	Сбор и хранение больших данных.	Поиск источников информации в сети Интернет: открытые и закрытые источники данных. Портал открытых данных РФ.
6	Методы обработки и анализа данных	Группировка данных, обнаружение значимых корреляций, зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного типа. Применение различных методов выделения, извлечения и группировки данных, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий
7	Визуализация исходной информации и аналитических данных.	Возможности графического представления информации: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (часов)				
		Лекции	Практические	Лабораторные	Самостоятельная работа	Всего
1	Введение в анализ больших данных. Обзор источников информации.	2	4		8	14
2	Технологии хранения и обработки больших данных.	2	4		8	14
3	Статистические методы анализа данных.	2	6		8	16
4	Современные программные средства анализа больших данных.	2	4		8	14
5	Сбор и хранение больших данных.	2	4		8	14
6	Методы обработки и анализа данных	4	6		10	20
7	Визуализация исходной информации и аналитических данных.	2	6		8	16
Итого:		16	34		58	108

14. Методические указания для обучающихся по освоению дисциплины

Для освоения дисциплины обучающимся необходимо работать с лекционными материалами (конспектами лекций) и практическими заданиями, размещенными на образовательном портале <https://edu.vsu.ru/>, основной и дополнительной литературой, выполнять задания на практических занятиях и в процессе самостоятельной работы, пройти текущие аттестации.

Дополнительные методические рекомендации по выполнению практических заданий, а также замечания по результатам их выполнения могут размещаться на портале <https://edu.vsu.ru/> в виде индивидуальных комментариев и файлов обратной связи, сообщений форума и других элементов электронного курса.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины (список литературы оформляется в соответствии с требованиями ГОСТ и используется общая сквозная нумерация для всех видов источников)

а) основная литература:

№ п/п	Источник
1	Адлер, Ю. Практическое руководство по статистическому управлению процессами / Ю. Адлер, В. Л. Шпер ; ред. В. Ионон. – Москва : Альпина Паблишер, 2019. – 234 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=570307 (дата обращения: 04.07.2022)
2	Москалев, С.М. Интернет-технологии и реклама в бизнесе: учебное пособие / С.М. Москалев ; Министерство сельского хозяйства Российской Федерации, Санкт-Петербургский государственный аграрный университет. – Санкт-Петербург : Санкт-Петербургский государственный аграрный университет (СПбГАУ), 2018. – 101 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=491717

б) дополнительная литература:

№ п/п	Источник
1.	Акперов И. Г. Информационные технологии в менеджменте: Учебник / И.Г. Акперов, А.В. Сметанин, И.А. Коноплева. - М.: ИНФРА-М, 2012. - 400 с.
2.	Балдин К.В., Уткин В.Б. Информационные системы в экономике : учебник .- Москва: Дашков и К-, 2012. - 396 с.

3.	Гладких, Т. В. Информационные системы учета и контроля ресурсов предприятия : учебное пособие : [16+] / Т. В. Гладких, Л. А. Коробова, М. Н. Ивлиев ; науч. ред. Д. С. Сайко ; Воронежский государственный университет инженерных технологий. – Воронеж : Воронежский государственный университет инженерных технологий, 2020. – 89 с. : ил., табл., схем., граф. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=612378 (дата обращения: 04.07.2022).
4.	Ж .Ж. Ламбен. Менеджмент, ориентированный на рынок: Питер, 2011. – 720 с.
5.	Мхитарян С.В. Маркетинговая информационная система: Учебно-методический комплекс. – М.: Изд. Центр ЕАОИ. 2008. 112 с.
6/	Туманов, В. Е. Проектирование хранилищ данных для систем бизнес-аналитики : учебное пособие / В. Е. Туманов. – Москва : Интернет-Университет Информационных Технологий (ИНТУИТ) : Бином. Лаборатория знаний, 2010. – 616 с. : ил., табл., схем. – (Основы информационных технологий). – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=233492 (дата обращения: 04.07.2022)

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

№ п/п	Ресурс
1.	Национальный цифровой ресурс «РУКОНТ» – <URL: http://rucont.ru/
2.	ЭБС Издаиельства «Лань» – <URL: http://www.e.lanbook.com/
3.	ЭБС «Университетская библиотека Online» – <URL: http://www.biblioclub.ru/
4.	Электронный университет https://edu.vsu.ru/

16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Конспекты лекций, размещенные на https://edu.vsu.ru/
2	Задания для практических занятий, размещенные на https://edu.vsu.ru/

17. Информационные технологии, используемые для реализации учебной дисциплины, включая программное обеспечение и информационно-справочные системы (при необходимости)

Дисциплина реализуется с применением элементов электронного обучения и дистанционных образовательных технологий (ЭОиДОТ) («Электронный университет» <https://edu.vsu.ru/>).

Используются такие средства информационно-коммуникационных технологий. как текстовые редакторы, электронные таблицы, средства подготовки презентаций, облачный сервис Яндекс.

Программный продукт Microsoft Office стандартный 2010, содержит Microsoft Word, IBM SPSS® Statistics Base, Microsoft Publisher, Microsoft PowerPoint. Версия 14.0.7128.5000.

17. . Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ, электронное обучение (ЭО), смешанное обучение):

Дисциплина реализуется с применением элементов электронного обучения и дистанционных образовательных технологий (ЭОиДОТ) («Электронный университет»).

18. Материально-техническое обеспечение дисциплины:

Компьютерный класс: 25 персональных компьютеров HP ProDesk 400 G5 DM/SATA 1Tb/Мо нитор ЖК 21,5" BenQ BL2283, 1920*1080 LED, 16:9, 250кд, 1000:1, DC 20000000:1, 5мс, IPS, 178/178, HDMI, колонки мультимедийный проектор NEC, экран настенный 153×200. Программное обеспечение: Office Standard 2019 Single OLV NL Each AcademicEdition Additional Product, Win Pro 10 32-bit/64-bit All Lng PK Lic Online DwnLd NR. Неисключительные права на ПО Dr. Web Enterprise Security Suite Комплексная защита Dr. Web Desktop Security Suite, браузер Google Chrome, WinRAR.

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1	Введение в анализ больших данных. Обзор источников информации.	ПК 1 Способен формировать возможные решения на основе разработанных для них целевых показателей	ПК 1.1 Способен осуществлять выявление, сбор и анализ информации бизнес-анализа для формирования возможных решений	Доклад
2	Технологии хранения и обработки больших данных.			Лабораторные задания
3	Статистические методы анализа данных.			Лабораторные задания
4	Современные программные средства анализа больших данных.			Лабораторные задания
5	Сбор и хранение больших данных.			Доклад
6	Методы обработки и анализа данных			Контрольная работа
7	Визуализация исходной информации и аналитических данных.			Доклад
8	Применение и кластерного анализа в специализированных программах			Лабораторные задания
9	Применение факторного анализа в специализированных программах			Лабораторные задания
Промежуточная аттестация, форма контроля – зачет				Перечень вопросов, пример КИМ приведены в п.20.2

20. Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1. Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

20.1.1 Перечень тем для докладов

1. Бизнес-проблемы и наука о данных. Формулировка бизнес-проблем. Решения, основанные на данных.
2. Проектная и процессная организация аналитики.
3. Business Intelligence. Business Analytics. Определение и соотношение понятий
4. Enterprise Decision Management. Суть концепции. Смысл управления решениями в организации
5. Data Science и Big Data. Определение и соотношение понятий с точки зрения бизнеса и инженерии
6. Базы данных и хранилища данных.
7. Функциональные классы аналитических систем.
8. Системы оптимизации. Экспертные системы. Системы машинного обучения.
9. Операционная бизнес-аналитика. Аналитическая отчетность. ERP-системы.
10. Облачные решения в области бизнес-аналитики
11. Инфраструктура анализа данных
12. Приложения и сервисы, основанные на данных

13. A/B тестирование и оптимизационные алгоритмы.

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Полное соответствие ответа обучающегося всем перечисленным критериям: полнота охвата выбранной темы; привлечение дополнительных источников литературы, в том числе из статей в ведущих журналах по теме, докладов ведущих конференций; структурирование презентации и управление временем доклада; уместное применение иллюстративного материала; корректность оформления ссылок; управление обсуждением доклада	<i>Базовый уровень</i>	<i>Зачтено</i>
Ответ на контрольно-измерительный материал не соответствует любым трем из перечисленных показателей: полнота охвата выбранной темы; структурирование презентации и управление временем доклада; уместное применение иллюстративного материала; корректность оформления ссылок; управление обсуждением доклада. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки, не владеет материалом	-	<i>Не зачтено</i>

20.1.2 Перечень лабораторных заданий

Подобрать данные для таблицы, приведенной ниже и проанализировать их взаимное влияние, отобразить корреляцию:

1. Роста ВВП на прирост населения
2. Прироста населения на динамику безработицы
3. Прирост людей с высшим образованием на рост промышленного производства
4. Прирост людей с высшим образованием на развитие науки
5. Прирост людей с высшим образованием на динамику доходов на душу населения
6. Динамику безработицы на динамику преступности
7. С помощью регрессионного анализа найдите зависимые переменные и поясните влияние на них независимых переменных.

Период	Численность населения	Рост ВВП	Рост ВВП на душу населения	Динамика промышленного производства	Развитие науки (высокотехнологичных отраслей)	Динамика доходов на душу населения

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Полное соответствие ответа обучающегося всем перечисленным критериям. Продемонстрировано знание принципов применения методов одномерного и многомерного статистического анализа, умения выбирать процедуры обработки информации в зависимости от природы используемых данных, навыками обработки экономической информации	<i>Базовый уровень</i>	<i>Зачтено</i>
Ответ на контрольно-измерительный материал не соответствует любым трем из перечисленных показателей, обучающийся дает неполные ответы на дополнительные вопросы. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки, не владеет основными методами статистического анализа, не демонстрирует умения обработки экономической информации	-	<i>Не зачтено</i>

20.1.3 Перечень заданий для контрольных работ

1. Найти потоковые данные, формирующие базу больших данных.
2. Провести процедуру структуризации и записи данных.
3. Написать функцию обращения к данным.
4. Провести визуализацию и первичный статистический анализ больших данных.

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Полное соответствие ответа обучающегося всем перечисленным критериям. Продемонстрировано знание принципов применения методов одномерного и многомерного статистического анализа, умения выбирать процедуры обработки информации в зависимости от природы используемых данных, навыками обработки экономической информации с применением специализированных программ	<i>Базовый уровень</i>	<i>Зачтено</i>
Ответ на контрольно-измерительный материал не соответствует любым трем из перечисленных показателей, обучающийся дает неполные ответы на дополнительные вопросы. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки, не владеет основными методами статистического анализа, не демонстрирует умения обработки экономической информации	-	<i>Не зачтено</i>

20.2. Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств:

Перечень вопросов к зачету:

1. Охарактеризуйте возможности применения программы для экономического анализа.
2. Интерпретация результатов факторного анализа
3. Какие существуют основные характеристики распределения, алгоритм вычисления
4. Модель факторного анализа, принципы выбора числа факторов
5. Алгоритм вычислений факторного анализа.
6. Определите сущность понятия «большие данные».
7. Основные вызовы больших данных.
8. Процесс аналитики анализа больших данных.
9. Дайте характеристику Big Data на мировом рынке.
10. Охарактеризуйте Big Data в России.
11. Определите понятие Data Mining.
12. В чем состоит когнитивный анализ данных.
13. Какие модели данных вы знаете?
14. Основные описательные статистики.
15. Определите различия между параметрическими, непараметрическими и номинальными методами.
16. Опишите основную идею корреляционного анализа.
17. Регрессионный анализ.
18. Основная идея дисперсионного анализа.
19. Сущность кластерного анализа.
20. Дискриминантный анализ: модель и общая процедура выполнения.
21. Цели факторного анализа.
22. Программные средства анализа данных: Statistica, SPSS, Excel и другие; их преимущества и недостатки.
23. Преимущества работа с данными в программе R.
24. Представление исходных данных в программе R.
25. Выполнение анализа данных в R.
26. Цели применения кластерного анализа
27. Условия применения параметрических критериев при проверке гипотез
28. Методы объединения кластеров (иерархический кластерный анализ)
29. Условия применения непараметрических критериев при проверке гипотез
30. Методы объединения кластеров (метод K-средних)
31. Основные типы шкал и соответствующие им меры средней тенденции и меры разброса
32. Алгоритм вычислений кластерного анализа
33. Способы проверки гипотез о соответствии эмпирического распределения одному из теоретических законов
34. Какие показатели характеризуют форму и тесноту корреляционной связи

35. Сущность дисперсионного анализа
36. Общее описание регрессионной модели. Особенности использования регрессионных моделей при анализе данных выборочных исследований
37. Задачи, решаемые с помощью дисперсионного анализа
38. Множественный регрессионный анализ. Проверка качества полученной модели, требования к исходным данным.
39. Алгоритм вычислений одномерного дисперсионного анализа
40. Цели и задачи корреляционного анализа при анализе экономических данных

Промежуточная аттестация проводится в соответствии с Положением о промежуточной аттестации обучающихся по программам высшего образования. Промежуточная аттестация обучающихся осуществляется в форме зачета.

Для оценивания результатов обучения на зачете используется – зачтено, не зачтено
Соотношение показателей, критериев и шкалы оценивания результатов обучения.

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Полное соответствие ответа обучающегося всем перечисленным критериям. Продемонстрировано знание принципов применения методов одномерного и многомерного статистического анализа, умения выбирать процедуры обработки информации в зависимости от природы используемых данных, навыками обработки экономической информации с применением специализированных программ	<i>Базовый уровень</i>	<i>Зачтено</i>
Ответ на контрольно-измерительный материал не соответствует любым трем из перечисленных показателей, обучающийся дает неполные ответы на дополнительные вопросы. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки, не владеет основными методами статистического анализа, не демонстрирует умения обработки экономической информации	-	<i>Не зачтено</i>

Программа рекомендована НМС факультета МО
протокол № от 15.03.2023