Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 31.05.2024 15:12:15
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
Факультет информационных технологий

УТВЕРЖДАЮ
Декан факультета
«Информационные технологии»
_____ / Д.Г.Демидов /
«15» февраля 2024г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Бизнес-аналитика и визуализация данных»

Направление подготовки/специальность
01.04.02 Прикладная математика и информатика

Профиль/специализация
«Системная аналитика больших данных»

Квалификация
магистр

Формы обучения
очная

Москва, 2024 г.

Разработчик(и):

к.э.н., доцент / С.В. Куликова /

Согласовано:
Заведующий кафедрой «Прикладная информатика»,
к.э.н., доцент / С.В. Суворов /

Содержание

Оглавление

1 Цели, задачи и планируемые результаты обучения по дисциплине 4
2 Место дисциплины в структуре образовательной программы 5
3 Структура и содержание дисциплины 5
3.1 Виды учебной работы и трудоемкость 5
3.2 Тематический план изучения дисциплины 6
3.3 Содержание дисциплины 6
3.4 Тематика семинарских/практических и лабораторных занятий 7
3.5 Тематика курсовых проектов (курсовых работ) 7
4 Учебно-методическое и информационное обеспечение 7
4.1 Нормативные документы и ГОСТы 7
4.2 Основная литература 8
4.3 Дополнительная литература 8
5 Лицензионное и свободно распространяемое программное обеспечение 9
5.1 Современные профессиональные базы данных и информационные справочные системы 9
5.2 Интернет-ресурсы 10
6 Материально-техническое обеспечение 11
7 Методические рекомендации 11
7.1 Методические рекомендации для преподавателя по организации обучения 11
7.2 Методические указания для обучающихся по освоению дисциплины 11
8 Фонд оценочных средств 12
8.1 Методы контроля и оценивания результатов обучения 12
8.2 Шкала и критерии оценивания результатов обучения 12
8.3 Оценочные средства 14
Типовые оценочные материалы по теме 1 14
Задание 4. Модуль pandas 15
Задание 5. Создание функций и классов 15
Тест 16
Типовые оценочные материалы по теме 2 Задания по теме 2 20
Типовые вопросы для опроса по теме 2: 21
Типовые оценочные материалы по теме 3Задания по теме 3 21
Типовые вопросы для опроса по теме 3: 24
Тест 24
Типовые оценочные материалы по теме 4 Задания по теме 4 29
Типовые вопросы для опроса по теме 4: 30
Экзаменационные вопросы 30
Типовые вопросы, выносимые на экзамен 32
Типовые задачи на экзамен 32
Описание системы оценивания 36

1 Цели, задачи и планируемые результаты обучения по дисциплине

К **основным целям** освоения дисциплины «Бизнес-аналитика и визуализация данных» следует отнести:

Знать:

- Методы анализа данных и машинного обучения;
- Возможности программных средств статистической обработки и визуализации данных на уровне умения:

Уметь:

- Применять программные средства анализа данных, поддержки принятия решений;
- Применять современные средства бизнес-аналитики,
- Разрабатывать интерактивные отчеты, dashboard, пользоваться ими при решении задач бизнес-анализа

Владеть:

- Современными инфокоммуникационными технологиями, средства и пакетами анализа данных и визуализации;

К **основным задачам** освоения дисциплины «Бизнес-аналитика и визуализация данных» следует отнести:

– освоение методологии математического моделирования экономических процессов;

– использование компьютерных технологий реализации методов математического моделирование экономических процессов.

Обучение по дисциплине «Бизнес-аналитика и визуализация данных» направлено на формирование у обучающихся следующих компетенций:

Код и наименование компетенций	Индикаторы достижения компетенции
УК-3. Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	**Знать:** методики формирования команд; методы эффективного руководства коллективами; основные теории лидерства и стили руководства; **Уметь:** разрабатывать план групповых и организационных коммуникаций при подготовке и выполнении проекта; сформулировать задачи членам команды для достижения поставленной цели; разрабатывать командную стратегию; применять эффективные стили руководства командой для достижения поставленной цели; **Владеть:** умением анализировать, проектировать и организовывать межличностные, групповые и организационные коммуникации в команде для достижения поставленной

	цели; методами организации и управления коллективом.
ОПК-3. Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	**Знать:** принципы, методы и средства анализа для разрабатывания математических моделей и их анализа при решении задач в области профессиональной деятельности; **Уметь:** разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности; **Владеть:** навыками разработки математических моделей и их анализ при решении задач в области профессиональной деятельности.

2 Место дисциплины в структуре образовательной программы

Дисциплина «Бизнес-аналитика и визуализация данных» относится к дисциплинам по выбору студента основной образовательной программы магистратуры.

Она взаимосвязана логически и содержательно-методически со следующими дисциплинами ООП:

- Научно-исследовательская и проектная деятельность;
– Непрерывные математические модели.

3 Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 4 зачетных единицы т.е. 144 академических часа 36 часов лекций, 36 часов лабораторные занятия и 78 часа – самостоятельная работа студентов, зачет - 2 семестр и экзамен - 3 семестр..

3.1 Виды учебной работы и трудоемкость (по формам обучения)

3.1.1 Очная форма обучения

№ п/п	Вид учебной работы	Количество часов	Семестры	
			2	3
1	**Аудиторные занятия**			
	В том числе:			
1.1	Лекции	22	12	10
1.2	Семинарские/практические занятия			
1.3	Лабораторные занятия	44	24	20
2	**Самостоятельная работа**			
	В том числе:			
2.1	Подготовка к практическим занятиям	78	36	42
3	**Промежуточная аттестация**			
	Зачет/диф.зачет/экзамен		зачет	экзамен
	Итого:	**144**	72	72

3.2 Тематический план изучения дисциплины (по формам обучения)

3.2.1 Очная форма обучения

№ п/п	Разделы/темы дисциплины	Трудоемкость, час					
		Всего	Аудиторная работа				Самостоятельная работа
			Лекции	Семинарские/ практические занятия	Лабораторные занятия	Практическая подготовка	
1	Раздел 1.						
1.1	Анализ данных. Построение графикови визуализация	36	6		12		18
1.2	Методы сетевого анализа	36	6		12		18
1.3	Агрегирование и проецирование	36	5		10		21
1.4	Средства бизнес-аналитики	36	5		10		21
Итого		**144**	**22**		**44**		**78**

3.3 Содержание дисциплины

Тема 1. Анализ данных. Построение графиков и визуализация

Понятие анализа данных. Data Mining. Стандарты Data Mining. Стандарт CWM, CRISP, PMML. Жизненный цикл процесса анализа данных. Управление на основе данных. Общая характеристика науки о данных. Языки разметки. Квартет Энскомба. 2D и 3D-диаграммы. Параметры диаграмм. Построение диаграмм в Python и R. Диаграммы рассеяния. Статистические диаграммы, гистограммы, функции ядерной плотности. Диаграммы размахов. Круговые и столбиковые диаграммы. Скрипичные диаграммы. Предобработка данных. Очистка данных. Борьба с пропусками и аномалиями. Описательная статистика и подгонка распределений. Классические методы и критерии статистики. Таблицы сопряженности. Номинальные, порядковые и интервальные статистики.

Методы повторной выборки. Бутстреп. Кросс-проверка. Перестановочный тест.

Метод складного ножа.

Тема 2. Методы сетевого анализа

Основы сетевого анализа. Основные понятия теории графов. Показатели графа. Размер. Плотность. Компоненты. Диаметр. Коэффициент кластеризации. Центральность. Социальная сеть. Пример социальной сети Морено. Меры центральности. Центральность по близости. Центральность по посредничеству. Кластеризация на графах. Построение графов в R. Библиотеки работы с графами и социальными сетями igraph, network. Графическое представление и укладка сети. Алгоритмы укладки. Дизайн сетей. Дуговые диаграммы, хордовые диаграммы. Тепловые карты. Социальная сплоченность. Клики. Обнаружение сообществ. Модулярность. Сети аффилированности. Динамические сети. Моделирование

динамических сетей. Модель случайного графа Эрдеша-Реньи. Программа Gephi, Orange. Графические виджеты.

Тема 3. Агрегирование и проецирование данных

Агрегирование и группирование данных. Операторы агрегирования. Сводная таблица. Power pivot. Power view. Понятие проекции. Принцип ординации наблюдений. Методы ординации. Основные многомерные методы статистического анализа. Задачи снижения размерности данных. Метод главных компонент. Факторный анализ. Основные этапы факторного анализа. Критерии факторного анализа. Матрица факторных нагрузок. Общность и характерность переменных. График факторных нагрузок. Методы вращения Многомерное шкалирование. Понятие стресса. Анализ соответствий.

Тема 4. Средства бизнес-аналитики

Общая характеристика средств бизнес-аналитики. Платформа Power BI. Основные задачи внедрения Power BI в компании. Состав платформы. Редактор Power BI querry Editor. Средства визуализации данных. Основные этапы составления отчета. Публикация отчета и совместная работа. Семейство MS Power BI. Конструктор кубов и аналитических представлений. Совместная работа с представлениями – серверные сервисы Power BI. Решение задач объединения данных из разных источников, вычисляемые поля и Меры, элементы языка DAX, построение различных визуальных представлений по данным из учетных систем и CRM–систем. Примеры построения приложений. Платформа Qlik sense. Особенности построения отчетов с помощью Qlik sense. Общая характеристика Tableau. Общая характеристика ИАС.

3.4 Тематика семинарских/практических и лабораторных занятий

Тема (раздел)	Формы текущего контроля успеваемости
Тема 1. Анализ данных. Построение графиков и визуализация	УО/Зад/Т
Тема 2. Методы сетевого анализа	УО/Зад
Тема 3. Агрегирование и проецирование данных	УО/Зад/Т
Тема 4. Средства бизнес-аналитики	УО/Зад

3.5 Тематика курсовых проектов (курсовых работ)

Не предусмотрено

4 Учебно-методическое и информационное обеспечение

4.1 Нормативные документы и ГОСТы

1. Федеральный государственный образовательный стандарт высшего образования (уровень магистратуры) по направлению подготовки 01.04.02 Прикладная математика и информатика, утвержденный приказом Министерства науки и высшего образования Российской Федерации от 10.01.2018 №13.

4.2 Основная литература

1. Афанасьев, Владимир Николаевич. Анализ временных рядов и прогнозирование. - Саратов:Ай Пи Ар Медиа. – 310 с. Текст электронный. - URL: http://www.iprbookshop.ru/90196.html (дата обращения: 12.11.2023). - Режим доступа: для авторизир. Пользователей.

2. Курносов М.Г. Введение в методы машинной обработки данных. - Новосибирск: Автограф. -220 с. Текст: электронный. - URL: http://www.iprbookshop.ru/102117.html (дата обращения: 11.01.2021). - Режим доступа: для авторизир. пользователей

Мастицкий С. Э. (2020) Анализ временных рядов с помощью R. — Электронная книга, адрес доступа: https://ranalytics.github.io/tsa-with-r

Миркин, Борис Григорьевич. Введение в анализ данных – М.:Юрайт, 2020 – 174 с. Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/450262 (дата обращения: 01.10.2023)

5 Мхитарян В. С., Архипова М. Ю., Дуброва Т. А., Миронкина Ю. Н., СиротинВ. П. Анализ данных. – М.: Юрайт, 2020 – 490 с. Текст : электронный // ЭБС Юрайт [сайт].

— URL: https://urait.ru/bcode/450166 (дата обращения: 29.09.2020)

6 О`Нил, Кэти. Data Science : Инсайдерская информация для новичков. Включаяязык R : [пер. с англ.] – СПб. Питер. – 368 с. Текст: электронный. - URL: http://new.ibooks.ru/bookshelf/359209/reading (дата обращения: 25.01.2023)

Сузи Р.А. Python. – СПб.: БХВ – Петербург, 2015 – 759 с. электронный ресурс Джеймс Г., Уиттон Д., Хасти Т., Тибширани Р. (2016) Введение в статистическое

обучение с примерами на языке R. - Пер. с англ. С. Э. Мастицкого. - М.: ДМК Пресс. - https://github.com/ranalytics/islr-ru

9. Кудрявцев, П. С. Анализ линейных детерминированных математических моделей технических объектов и синтез линейных систем управления : учебное пособие / П. С. Кудрявцев. — Москва : МАИ, 2022. — 91 с. — ISBN 978-5-4316-0981-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/344042 (дата обращения: 27.09.2023). — Режим доступа: для авториз. пользователей.

10 Семенов, А. Д. Математические модели систем управления : учебное пособие / А. Д. Семенов, А. В. Волков, О. В. Ермилина. — Вологда : Инфра-Инженерия, 2022. — 200 с. — ISBN 978-5-9729-0889-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/281216 (дата обращения: 27.09.2023). — Режим доступа: для авториз. пользователей.

4.3 Дополнительная литература

Гусарова Н.Ф, Анализ социальных сетей. Основные понятия и метрики. – СПб.:Университет ИТМО, 2016 – 67 с.

Жерон, Орельен. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow : Концепции, инструменты и техники для создания интеллектуальных систем : полноцветное издание : перевод с английского - ПрМ.:Диалектика. -684 с.

Ланц Б. – Машинное обучение на R/ - СПб. :Питер. – 2020 – 464 с.

Principles of Econometrics with R [Электронный ресурс] – URL: https://bookdown.org/ccolonescu/RPoE4/

13. Люк Д. Анализ сетей (графов) в среде R. Руководство пользователя- М.: ДМК Пресс,2017 – 250 с.

Маккинни У. Python и анализ данных. – М.: ДМК, 2015 – 482 с.

Нильсен Эйлин. Практический анализ временных рядов: прогнозирование со статистикой и машинное обучение. –М.: ООО Диалектика – 2021 – 544 с.

Наумов, Владимир Николаевич. Средства бизнес- аналитики: учеб. пособие / В. Н. Наумов ; Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2016. - 107 с.

Наумов В.Н. Анализ данных и машинное обучение: методы и инструментальные средства. Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2020. - 260 с.

Шолле Ф. Глубокое обучение на Python. – СПб.: Питер. 2018. -400 с.

Шолле Ф. Глубокое обучение на R. – СПб.: Питер. 2018. -400 с.

11 Задорожная, Н. М. Основы теории и проектирования систем управления. Методология. Математические модели : методические указания / Н. М. Задорожная. — Москва : МГТУ им. Н.Э. Баумана, 2016. — 36 с. — ISBN 978-5-7038-4306-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/103603 (дата обращения: 27.09.2023). — Режим доступа: для авториз. пользователей

5 Электронные образовательные ресурсы

https://online.mospolytech.ru/course/view.php?id=3339 - Математическое моделирование экономических процессов

5 Лицензионное и свободно распространяемое программное обеспечение

1. Microsoft Windows.
2. Microsoft Visio.
3. Microsoft Office.
4. Power BI - https://powerbi.microsoft.com/ru-ru/
5 Аналитическая платформа Loginom - https://loginom.ru/
6 Техническая документация Майкрософт - https://docs.microsoft.com/ru-ru/

5.1 Современные профессиональные базы данных и информационные справочные системы

1. 1http://www.raai.org/ Российская ассоциация искусственного интеллекта
2. http://fuzzy.raai.org/ Российская ассоциация нечетких систем
3. http://aihandbook.intsys.org.ru/index.php/resources/orgs/242-org-p12 Ассоциация нейроинформатики/RNNS
4. www.aaai.org Американская ассоциация искусственного интеллекта American Association for Artificial Intelligence (AAAI)
5. http://aifuture.chat.ru/ Искусственный интеллект ("Взгляд в будущее").
6. http://aicommunity.narod.ru/ Материалы об искусственном интеллект.
7. Онлайн курс «Машинное обучение» - https://www.coursera.org/learn/machine-learning
8. Онлайн курс «Нейронные сети и компьютерное зрение» - https://stepik.org/course/50352
9. Онлайн курс ШАД «Машинное обучение» - https://yandexdataschool.ru/edu-process/courses/machine-learning

10. Информационный ресурс www.machinelearning.ru.

5.2 Интернет-ресурсы

1. Электронно-образовательные ресурсы на сайте научной библиотеки СЗИУ РАНХиГС(http://nwipa.ru)
2. Электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс»http://www.nwapa.spb.ru/index.php?page_id=76
3. Электронные учебники электронно-библиотечной системы (ЭБС) «Лань»http://www.nwapa.spb.ru/index.php?page_id=76
4. Электронные учебники электронно-библиотечной системы (ЭБС) «IPRbooks»http://www.nwapa.spb.ru/index.php?page_id=76
5. Электронные учебники электронно-библиотечной системы (ЭБС) «Юрайт»
6. http://www.nwapa.spb.ru/index.php?page_id=76
7. Научно-практические статьи по экономике и финансам Электронной библиотеки ИД«Гребенников» http://www.nwapa.spb.ru/index.php?page_id=76
8. Статьи из журналов и статистических изданий Ист-Вью http://www.nwapa.spb.ru/index.php?page_id=76
9. Англоязычные ресурсы EBSCO Publishing: доступ к мультидисциплинарным полнотекстовым базам данных различных мировых издательств по бизнесу, экономике, финансам, бухгалтерскому учету, гуманитарным и естественным областям знаний,рефератам и полным текстам публикаций из научных и научно–популярных журналов.

Emerald eJournals Premier - крупнейшее мировое издательство, специализирующееся на электронных журналах и базах данных по экономике и менеджменту.

Возможно использование, кроме вышеперечисленных ресурсов, и других электронных ресурсов сети Интернет.

1. www.finexpert.ru
2. http://www.itnews.ru/
3. http://www.cnews.ru/
4. http://www.prj-exp.ru/
5. http://piter-consult.ru/
6. http://www.gartner.com /
7. http://www.idc.com
8. http://bpms.ru / BPMS.ru
9. http://www.betec.ru /
10. http://www.cfin.ru / Интернет-проект «Корпоративный менеджмент»
11. http://www.osp.ru / Открытые системы
12. http://www.citforum.ru / CIT forum
13. http://www.iteam.ru / Портал iTeam – Технологии корпоративного управления
14. http://www.idef.com / Методологии IDEF
15. http://www.interface.ru/home.asp?artId=4449 / Электронная версия книги Дэвид А. Марка, Клемент МакГоуэн Методология структурного анализа и проектирования SADT.
16. http://www.fa.ru/dep/cko/msq/Pages/default.aspx / Международные стандартыкачества.
17. http://office.microsoft.com/ru-ru/support/FX100996114.aspx / Microsoft Visio
18. http://www.businessstudio.ru / Business Studio
19. http://www.casewise.com/ru/products/corporate_modeler_suite.html / CasewiseCorporate Modeler Suite
20. http://www.interface.ru/home.asp?artId=106 / Process Modeler

21. http://www.oracle.com/technetwork/ru/middleware/bpa/index.html / Oracle BusinessProcess Analysis Suite
22. http://www.softwareag.com/ru / Software AG
23. http://www-03.ibm.com/software/products/ru/ru/wbi / IBM WebSphere BusinessModeler

6 Материально-техническое обеспечение

Четыре компьютерных класса Ауд. АВ4805, АВ4809, АВ4810, АВ4811, оснащенные методическими материалами по дисциплине (лекции, практические задания).

7 Методические рекомендации

7.1 Методические рекомендации для преподавателя по организации обучения

. 1. При подготовке к занятиям следует предварительно проработать материал занятия, предусмотрев его подачу точно в отведенное для этого время занятия. Следует подготовить необходимые материалы – теоретические сведения, задания и др. При проведении занятия следует контролировать подачу материала и решение заданий с учетом учебного времени, отведенного для занятия.

2. При проверке лабораторных работ и отчетов следует учитывать не только правильность выполнения заданий, но и оптимальность выбранных методов решения, правильность выполнения всех его шагов.

3. Текущий и промежуточный контроль проходит с учетом балла-рейтинговой системы оценивания.

7.2 Методические указания для обучающихся по освоению дисциплины

. Методические указания могут быть составлены для различных видов учебной работы, в том числе аудиторной, самостоятельной, подготовки к текущему контролю и промежуточной аттестации.

Изучение дисциплины осуществляется в строгом соответствии с целевой установкой в тесной взаимосвязи учебным планом. Основой теоретической подготовки студентов являются лекции и аудиторные занятия, лабораторные работы.

В процессе самостоятельной работы студенты закрепляют и углубляют знания, полученные во время аудиторных занятий, дорабатывают конспекты и записи, готовятся к промежуточной аттестации, а также самостоятельно изучают отдельные темы учебной программы.

На занятиях студентов, в том числе предполагающих практическую деятельность, осуществляется закрепление полученных, в том числе и в процессе самостоятельной работы, знаний. Особое внимание обращается на развитие умений и навыков установления связи положений теории с профессиональной деятельностью будущего специалиста.

Самостоятельная работа осуществляется индивидуально. Контроль самостоятельной работы организуется в двух формах:

• самоконтроль и самооценка студента;

• контроль со стороны преподавателей (текущий и промежуточный).

Текущий контроль осуществляется на аудиторных занятиях, промежуточный контроль осуществляется на зачете или экзамене в письменной (устной) форме.

Критериями оценки результатов самостоятельной работы студента являются:

• уровень освоения студентом учебного материала;

• умения студента использовать теоретические знания при выполнении практических задач;

• сформированность компетенций;

• оформление материала в соответствии с требованиями.

8 Фонд оценочных средств

8.1 Методы контроля и оценивания результатов обучения

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

- выполнение лабораторных работ;
- Зачет
- экзамен.

8.2 Шкала и критерии оценивания результатов обучения

Показателем оценивания компетенций на различных этапах их формирования является достижение обучающимися планируемых результатов обучения по дисциплине.

Показатель	**Критерии оценивания**			
	2	**3**	**4**	**5**
ОПК-3. Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности				
ОПК-1.1. Знать: принципы, методы и средства анализа для разрабатывания математических моделей и их анализа при решении задач в области профессиональной деятельности. ОПК-1.2. Уметь: разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности.	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие материалу дисциплины знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3).	Обучающийся демонстрирует неполное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает	Обучающийся демонстрирует частичное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Свободно оперирует приобретенными знаниями.

ОПК-1.3.Владеть: навыками разработки математических моделей и их анализ при решении задач в области профессиональной деятельности.		значительные затруднения при оперировании знаниями при их переносе на новые ситуации.		

Форма промежуточной аттестации: зачет.

Промежуточная аттестация обучающихся в форме зачёта проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «зачтено» или «не зачтено».

Шкала оценивания	**Описание**
Зачтено	Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
Не зачтено	Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

Форма промежуточной аттестации: экзамен.

Промежуточная аттестация обучающихся в форме экзамена проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов

обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «отлично», «хорошо», «удовлетворительно» или «неудовлетворительно».

Шкала оценивания	***Описание***
Отлично	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.*
Хорошо	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует неполное, правильное соответствие знаний, умений, навыков приведенным в таблицах показателей, либо если при этом были допущены 2-3 несущественные ошибки.*
Удовлетворительно	*Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, в котором освещена основная, наиболее важная часть материала, но при этом допущена одна значительная ошибка или неточность.*
Неудовлетворительно	*Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.*

8.3 Оценочные средства

Типовые оценочные материалы по теме 1

Задания по теме 1

Задание 1. Исследовать наборы данных в пакете ggplot2. Воспользоватьсяшпаргалками для решения задач построения диаграмм различных видов.

Задание2. Воспользоваться графическими приложениями RCmdr, Rattle для решениязадач графического анализа

Задание 3. Исследовать возможности графического пакета matplotlib, seaborn, Дляисследования воспользоваться блокнотом notebook

Задание 4. Сгенерировать последовательность равномерно распределенных случайных чисел, в диапазоне от 0 до 10. Размер последовательности 100. Аналогично сгенерировать еще одну последовательность целых чисел, в диапазоне от 0 до 20 также состоящую из 100 чисел. Используя функцию linspace() сгенерировать последовательностьиз 100 числе, заключенную от 0 до 10. Используя оператор цикла for и итератор i сформировать последовательность из 100 чисел от -1 до 9. Построить точечные, линейныедиаграммы для полученных последовательностей.

Построить матричную диаграмму, используя стиль matlab. Задавать панели с помощью subplot. Применить средства управления цветом, типом линии, маркерами.

Задание 4. Модуль pandas

Загрузить набор данных. Файл данных хранится в moodle.

```
sp500=pd.read_csv(filepath_or_buffer="спецификация файла",sep=',',
                  usecols=['Symbol','Sector','Price','B
    ook Value'],index_col='Symbol')
```

```
sp500.head()
```

Провести анализ файла.

С помощью метода iloc отобразить только 10,20,30,40,50, 60 и 70 элементыОтобразить только первые два столбца набора данных

С помощью метода sample выбрать 20 кампаний.

Задание 5. Создание функций и классов

Создайте функцию, которая решает систему алгебраических уравнений методов обратной функции, если определитель матрицы не равен нулю. В противном случае выдается сообщение о том, что система уравнений не определена. Проверить решение на нескольких вариантах исходных данных. Сравнить с результатами решения, полученнымивручную.

Постройте иерархию классов «Автомобиль». В иерархию классов включить классы: грузовые автомобили; легковые автомобили, гоночные автомобили, транспорт. В класс легковых автомобилей включить классы: седан, лимузин, купе, хетчбек, универсал, фургон, микроавтобус. Задать атрибуты «модель», «год выпуска». Задать свойства и методы для седана. Создать экземпляры седана

Типовые вопросы для опроса по теме 1

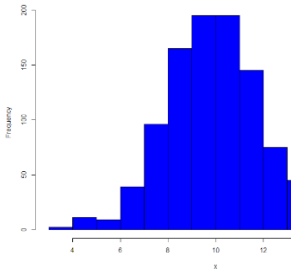
1. Назовите этапы анализа данных.
2. Для чего используется этап предобработки, а также этап разведочного анализа?
3. Выполните классификацию диаграмм. Приведите примеры использованиядиаграмм для решения различных задач графического анализа.
4. Какие задачи решаются во время предобработки данных. Назовите статистическиегипотезы, которые могут проверяться при решении задач разведочного анализа.
5. Дайте общую характеристику программных платформ работы с графикой.
6. Дайте характеристику графической грамматики. Приведите примеры е еиспользования для построения различных диаграмм
7. Назовите статистические диаграммы. Что представляет собой диаграммаКливленда?
8. Что такое повторная выборка? Какие методы применяются при решении задач повторной выборки?

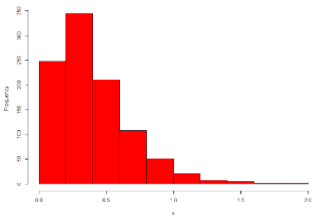
Тест

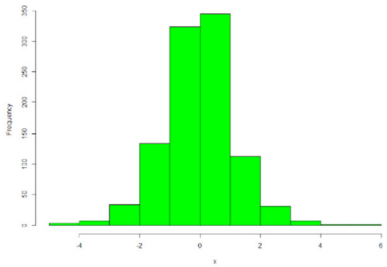
1. Гистограммы распределения

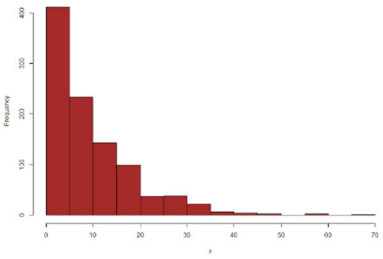
Поставить соответствие между заданными последовательностями и гистограммами распределения, если последовательности представляют собой случайные числа, полученные спомощью различных генераторов с разными законами распределения:

```
set.seed(1234)
x1<-
rnorm(1000,10,
2) x2<-
rgamma(1000,2,
5)
x3<-rt(1000,8)
x4<-rexp(1000,0.1)
```









Варианты ответов:

1. X1
2. X2
3. X3
4. X4

2. Критерий проверки гипотезы о нормальном законе распределения

Выберите из предложенного списка статистических критериев критерий, который позволяетпроверить гипотезу о нормальном законе распределения

- Колмогорова-Смирнова
- Пирсона (Хи квадрат)
- Стьюдента
- Фишера
- Шапиро
- Шапиро-Пятецкого
- Манна-Уитни
- Андерсона-Дарлинга

3. Асимметрия

Последовательность задана с помощью предложения

x<-c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)

Чему равна асимметрия данной последовательности, если все пропуски заменены средним

значением имеемых данных. Ответ дать с точностью до двух знаков после запятой

4. Значимость коэффициента корреляции

Чему равен уровень значимости при принятии гипотезы о наличии корреляции Пирсона длядвух случайных величин/ Ответ дать с точностью до двух знаков после запятой

Табли

ца
наблю
дений

x	y
0	3
2	4
3	5
1	3
4	4
5	3

5. Интерквартильный размах

Последовательность задана с помощью предложения

x<-c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)

Чему равен интерквартильный размах для данной последовательности, если все пропускизаменены средним значением имеемых данных. Ответ дать с точностью до целых

6. Корреляционный анализ

Чему равен коэффициент корреляции Спирмена для случайных величин X, Y

Таблица наблюдений

x	y
0	3
2	4
3	5
1	3
4	4
5	3

Ответ дать с точностью до двух знаков после запятой

7. коэффициент корреляции

Чему равен коэффициент корреляции Пирсона для случайных величинТаблица наблюдений

x	y
0	3
2	4
3	5
1	3
4	4
5	3

8. Критерий Шапиро-Уилка

Сгенерировать случайную выборку, размером в 500 наблюдений, используя генератор нормально распределенных случайных чисел с математическим ожиданием 10, дисперсией 16. При этом начальное значение генератора случайных чисел равно 4321. Чему равен уровень значимости при проверке гипотезы о нормальном законе распределения с помощью критерия Шапиро-Уилка. Ответ дать с точностью до двух знаков после запятой. В качестве альтернативной использовать двустороннюю гипотезу.

9. Проверка гипотезы о законе распределения

Сгенерировать случайную выборку, размером в 500 наблюдений, используя генератор нормально распределенных случайных чисел с математическим ожиданием 10, дисперсией 16. При этом начальное значение генератора случайных чисел равно 4321. Чему равно наблюдаемое значение критерия Колмогорова-Смирнова. Ответ дать с точностью до двух знаков после запятой. В качестве альтернативной использовать двустороннюю гипотезу.

10. Проверка гипотезы о значимости отличий

Имеется выборка значений двух случайных величин

x	y
0	3
2	4
3	5
1	3
4	4
5	3

Проверить гипотезу о равенстве математических ожиданий при допущении о неравных дисперсии при допущении о независимости данных величин.

Чему равно наблюдаемое значение критерия Стьюдента. Ответ дать с точностью до одного знака после запятой.

11. Проверка гипотезы о равенстве дисперсииДанные заданы двумя последовательностями x<-c(0,2,3,1,4,5)
y<-c(3,4,5,3,4,3)

Чему равно наблюдаемое значение критерия Фишера при проверке гипотезы о равенстве дисперсий? Ответ дать с точностью до двух знаков после запятой

12. Пропуски данных
Последовательность задана с помощью предложения

x<-c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)

Чему равно среднее данной последовательности, если все пропуски заменены медианным значением. Ответ дать с точностью до двух знаков после запятой

13. Пропуски данных
Последовательность задана с помощью предложения

x<-c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)

Чему равна медиана данной последовательности, если все пропуски заменены средним значением имеемых данных. Ответ дать с точностью до двух знаков после запятой

Ключи:

1	2	3	4	5	6	7	8	9	10	11	12	13
1-3,2-1,3-4, 4-2	1,4.5,7	0,86	0,62	9	0,28	0,26	0,57	0,96	-1,4	0,66	11,52	13,33

Типовые оценочные материалы по теме

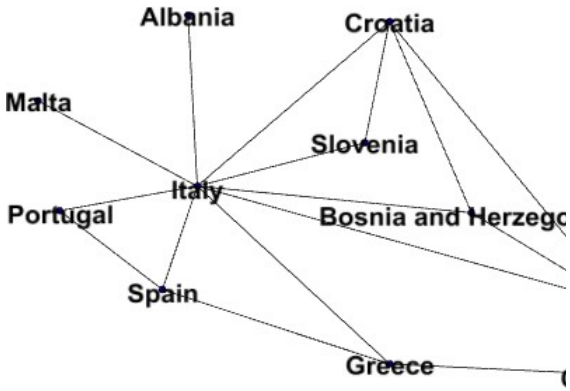
2 Задания по теме 2

Задаиие 1. Построить сетевой граф товарооборота между странами Северной Европы, если матрица смежности имеет вид:

	Norway	Lithuania	Latvia	Denmark	Sweden	Ireland	Finland	Estonia
Norway	0	1	0	1	0	0	1	1
Lithuania	1	0	0	0	1	0	0	0
Latvia	0	0	0	0	1	0	0	0
Denmark	1	0	0	0	0	0	1	0
Sweden	0	1	1	0	0	1	0	1
Ireland	0	0	0	0	1	0	0	0
Finland	1	0	0	1	0	0	0	1

Estonia	1	0	0	0	1	0	1	0

Задание 2. Решить задачу сетевого анализа графа, описывающие плотность торгово-экономических связей между странами Южной Европы, если граф имеет вид:



Задание 3. Построить граф аффилированности для файла данных UserNetR::hwd. В данном файле актеры Голливуда связаны между собой фильмами, в которых они снялись вместе.

Задание 4. Решить задачу построения и анализа графа из данных об оптовой торговле вином. Таблицу данных взять из сайта книги Джона Формана. Много цифр: Анвлиз больших данных при помощи Excel. – М.: Альпина Паблишер, 2016 – 461 с.

www.wiley.com/go/datasmart. Задачу решить в Gephi Исследовать модулярность графа. Построить кластеры модулярности.

Типовые вопросы для опроса по теме 2:

1. Дайте определение сети. Назовите способы задания сетей.
2. В чем состоят особенности социальных сетей?
3. Какие параметры рассчитываются при анализе сетей?
4. Показатели центральности сети?
5. Кластеризация на графах. Проведите сравнительный анализ методов, которые используются при кластеризации на графах
6. Что такое сети аффилированности? Приведите примеры.

Типовые оценочные материалы по теме

3Задания по теме 3

Задание 1. Решить задачу факторного анализа для набора данных ausland.sav. Данные представляют собой результаты анкетирования, проведенного Институтом Социологии Университета Марбург. На основе разработанной анкеты на двух гессенских металлургических предприятиях было произведено исследование

отношения к иностранцам. Опрашиваемым предложили высказать свое отношение к следующим пятнадцати положениям:

- 1. Необходимо улучшить интеграцию иностранцев.
- 2. Необходимо мягче относиться к беженцам.
- 3. Деньги Германии должны быть потрачены на нужды страны.
- 4. Германия — это не служба социальной помощи для всего мира.
- 5. Необходимо стараться налаживать хорошие отношения друг с другом.
- 6. Права беженцев следует ограничить.
- 7. Немцы станут меньшинством.
- 8. Право беженцев необходимо охранять во всей Европе.
- 9. Враждебность к иностранцам наносит вред экономике Германии.
- 10. Сначала необходимо создать нормальные жилищные условия для немцев.
- 11. Мы ведь тоже практически везде являемся иностранцами.
- 12. Мультикультура означает мультикриминал.
- 13. В лодке нет свободных мест.
- 14. Иностранцы вон.
- 15. Интеграция иностранцев — это убийство нации.

Оценки ставились по семибальной шкале: от полного несогласия (1) до полного согласия (7). Результаты опроса для 90 человек хранятся в файле ausland.sav в переменных *a1 - a15*. Пример приведен в самоучителе SPSS.

Отметим, что данные приведены на немецком языке. При выполнении задания следует данные представить на русском языке.

Задание 2. Решить задачу многомерного шкалирования, если матрица различи по восьми странам имеет вид

№п/п	Ангола	Аргентина	Австралия	Китай	Куба	Япония	США	Зимбабве
Ангола	0,00	1,41	1,00	1,00	1,41	1,41	1,73	0,71
Аргентина	1,41	0,00	1,00	1,73	1,41	1,41	1,00	1,41
Австралия	1,00	1,00	0,00	1,41	1,73	1,00	1,41	1,00
Китай	1,00	1,73	1,41	0,00	1,00	1,00	1,41	1,00
Куба	1,41	1,41	1,73	1,00	0,00	1,41	1,00	1,41
Япония	1,41	1,41	1,00	1,00	1,41	0,00	1,00	1,41
США	1,73	1,00	1,41	1,41	1,00	1,00	0,00	1,73
Зимбабве	0,71	1,41	1,00	1,00	1,41	1,41	1,73	0,00

Задание 3.Решить корреспондентного анализа для набора данных . smoking. Набор данных включает сведения о сотрудниках (старшие и младшие менеджеры, старшие и младшие сотрудники) и об их предрасположенности к курению (не курит, слабо, средне, сильно).

	Simple correspondence analysis example data file.			
	NONE	LIGHT	MEDIUM	HEAVY
SR.MANAGERS	4	2	3	2
JR.MANAGERS	4	3	7	4
SR.EMPLOYEES	25	10	12	4
JR.EMPLOYEES	18	24	33	13
SECRETARIES	10	6	7	2

Решить задачу корреспондентного анализа для набора данных

Должность	Курение	Алкоголь
Старший менеджер	нет	нет
Младший менежжер	слабо	да
Старший сотрудник	средне	иногда
Младший сотрудник	сильно	нет
Секретарь	нет	да
Старший менеджер	нет	нет
Младший менежжер	слабо	нет
Старший сотрудник	нет	да
Младший сотрудник	слабо	иногда
Секретарь	сильно	да
Старший менеджер	слабо	нет
Младший менежжер	нет	нет
Старший сотрудник	нет	да
Младший сотрудник	сильно	иногда
Секретарь	средне	иногда
Старший менеджер	средне	нет
Младший менежжер	нет	да
Старший сотрудник	нет	нет
Младший сотрудник	слабо	нет
Секретарь	нет	нет
Старший сотрудник	сильно	да
Младший сотрудник	нет	нет
Старший сотрудник	сильно	нет
Программист	средне	да
Программист	средне	нет
Программист	нет	иногда
Руководитель	нет	иногда
Руководитель	средне	нет
Руководитель	нет	нет
Программист	сильно	да
Секретарь	нет	нет

Младший сотрудник	средне	нет
Младший менежжер	нет	да

Типовые вопросы для опроса по теме 3:

1. В чем суть задачи многомерного шкалирования?
2. Как решается задача метрического шкалирования по Тор- герсону?
3. В чем отличие метрического шкалирования от неметрического?
4. Как строится матрица различий объектов?
5. Каков алгоритм решения задачи неметрического шкалирования?
6. Какие допущения принимаются при решении задач факторного анализа?
7. Назовите этапы факторного анализа?
8. Какие критерии используются при задании параметров факторного анализа?
9. Дайте общую характеристику задач ординации
10. Что такое корреспондентный анализ?

Тест

1. Этапы факторного анализа

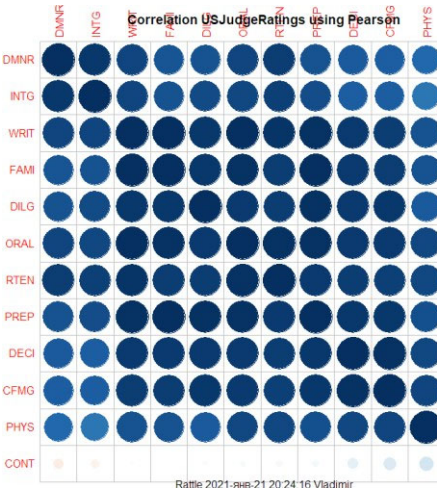
Поставьте соответствие между номерами этапов факторного анализа и их

содержанием

1. Выбор метода факторного анализа
2. Выбор числа компонент (факторов)
3. Построение матриц и определение значений компонент (факторов) для каждого наблюдения
4. Вращение компонент (факторов
5. Интерпретация полученных компонент (факторов)

2. Корреляционная матрица

Диаграмма корреляционной матрицы Пирсона имеет вид



Выберите правильный ответ

1. эффективность факторного анализа невысока
2. следует выполнять факторный анализ. Однако до этого необходимо

выбратьподмножество анализируемых переменных

3. Целесообразно решить задачу факторного анализа

4. Для принятия решения о целесообразности факторного анализа необходимоиспользовать непараметрические критерии корреляции

5. Факторный анализ не целесообразен. Нужно попробовать решить задачуметодом главных компонент

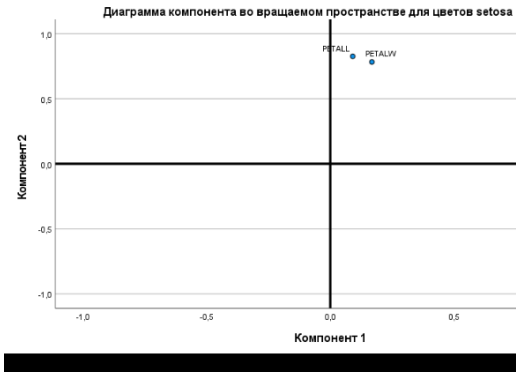
3. Матрица факторных нагрузок

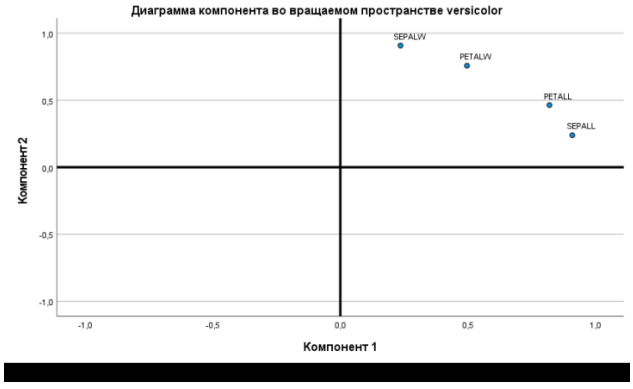
Что означает элемент матрицы факторных нагрузок aij?

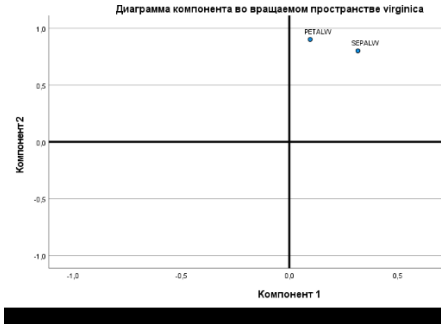
1. корреляцию между переменной и фактором (компонентом)
2. ковариацию между наблюдениями
3. суммарную дисперсию наблюдений для каждой переменных
4. дисперсию наблюдений для каждого фактора (компонента)
5. стандартизированное значение переменной

4. Оценка качества факторного анализа

В результате решения задачи факторного анализа получены графики нагрузок длякаждого типа цветка набора данных iris





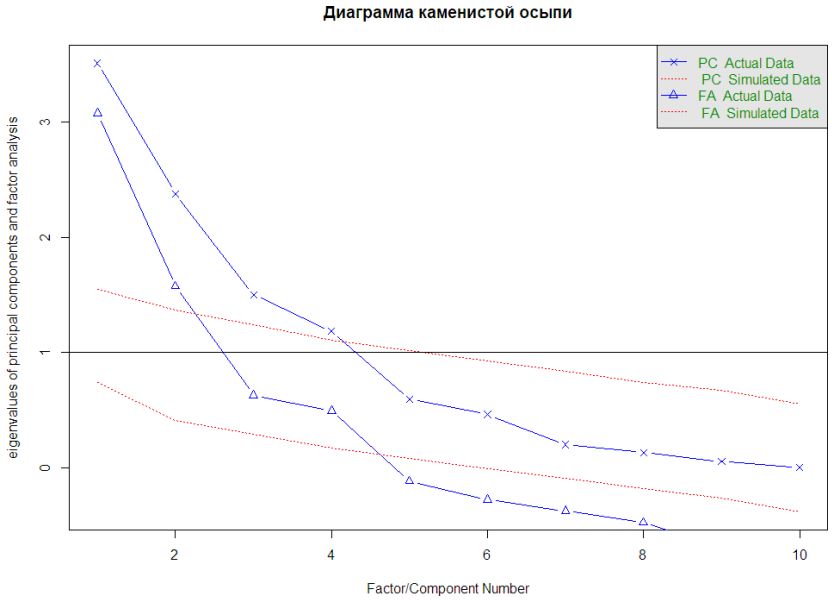


Выберите правильное ранжирование качества решения задачи факторного анализадля разных классов цветов

1. качество одинаковое
2. 1 setosa, 2. versicolor, 3 – virginica
3. 1 setosa, 2. - virginica, 3. - versicolor
4. 3 setosa, 2. - virginica, 1. - versicolor
5. 3 setosa, 1. - virginica, 2. - versicolor

5. Критерий Кайзера

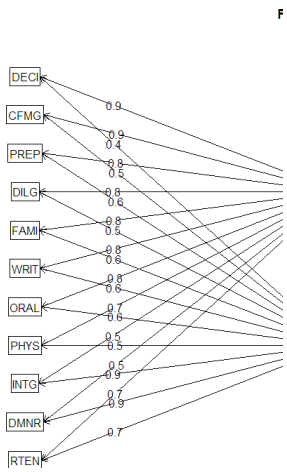
Диаграмма каменистой осыпи имеет вид



Чему равно рекомендуемое число главных компонент для фактических данных, еслииспользуется метод главных компонент? Ответ дать числом

6. Матрица факторных нагрузок

Исходный набор данных содержит 20 наблюдений и 11 переменных (feature). После анализа значений критериев факторного анализа принято решение о двух



факторах. Диаграмма результатов факторного анализа имеет вид:

Выберите правильную размерность матрицы факторных нагрузок, если превое числопары чисел определяет число строк, а второе - число столбцов

7. Метод главных компонент

В результате решения задачи сокращения размерности для набора iris методом главных компонент получены таблицы

```
Note that principal components on only the numeric
 variables is calculated, and so we can not use this
 approach to remove categoric variables from  consideration.

 Any numeric variables with relatively large rotation
 values (negative or positive) in any of the first few
 components are generally variables that you may wish
 to include in the modelling.

Rattle timestamp: 2021-01-21 12:37:54 Vladimir
========================================================================
Call:
princomp(x = na.omit(crs$dataset[crs$train, crs$numeric]), scale = TRUE,
    center = TRUE, tol = 0)

Standard deviations:
Comp.1 Comp.2 Comp.3 Comp.4
  2.02   0.52   0.30   0.16

 4  variables and  105 observations.

Rattle timestamp: 2021-01-21 12:37:54 Vladimir
========================================================================
Importance of components:
                       Comp.1 Comp.2 Comp.3 Comp.4
Standard deviation       2.02   0.52   0.30 0.1554
Proportion of Variance   0.92   0.06   0.02 0.0054
Cumulative Proportion    0.92   0.97   0.99 1.0000

Rattle timestamp: 2021-01-21 12:37:54 Vladimir
========================================================================
```

Какова доля объясненной дисперсии переменных первой главной компонентой? Ответ дать с точностью до двух знаков после запятой. В качестве разделителя использовать запятую

8. Метод главных компонент

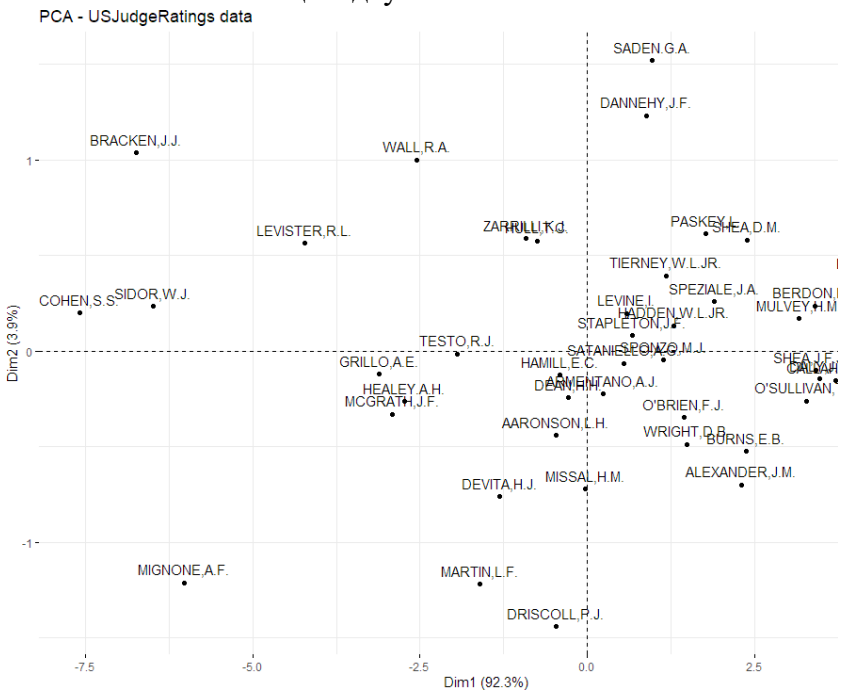
Для решения задачи факторного анализа в R выполнен следующий скрипт

library(psych)

pc <- principal(USJudgeRatings[,-1], nfactors=6,rotate='none'). Сколько главныхкомпонент достаточно для решения задачи факторного анализа? Ответ дать числом

9. Метод главных компонент

При решении задачи ординализации получена диаграмма. Какая доля дисперсии переменных описывается с помощью двух главных компонент?



Ответ дать с точностью до двух знаков после запятой

10. Собственные значения

При решении задачи факторного анализа методом главных компонент в SPSS получена таблица объясненной совокупной дисперсии.

Объясненная совокупная дисперсия

	Начальные собственные значения			Извлечение суммы квадратов наг		
Компонент	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Сумм
1	2,911	72,770	72,770	2,911	72,770	
2	,921	23,031	95,801	,921	23,031	
3	,147	3,684	99,485			
4	,021	,515	100,000			

Метод выделения факторов: метод главных компонент.

Чему равно собственное значение для первой главной компоненты? Ответ дать сточностью до двух знаков после запятой

Ключи:

1	2	3	4	5	6	7	8	9	10
1-1, 2-2, 3-3, 4-4, 5-5	2	1	3	4	(11,2)	0,92	1	0,96	2,91

Типовые оценочные материалы по теме

4 Задания по теме 4

Задание 1.

Использование пакета QlikView, Tableau и Power BI для решения задач анализа данных о демографической ситуации в России. Для каждого варианта приведены таблицы с указанием вида исходных данных, которые будут анализироваться средствами бизнес- аналитики.

Вариант	год	область	регион	городское население	зарплата	миграция	Млад. Смертность	Рождаемость	Смертность	население	осн. Фонды	преступления	сельскеХоз	Трудоспособное Население	безработные
1	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+
2	+	+	+	-	+	+	-	+	-	+	+	-	-	+	+
3	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+
4	+	+	+	-	+	-	+	-	-	+	+	-	+	+	-
5	+	+	+	+	-	+	-	-	+	+	-	+	-	+	-
6	+	+	+	-	+	+	-	-	-	+	+	-	-	+	-
7	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+
8	+	+	+	-	+	-	+	+	-	+	+	-	+	+	+
9	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+
10	+	+	+	-	+	+	-	-	-	+	+	-	-	+	-
11	+	+	+	+	-	-	+	-	+	+	-	+	+	+	-
12	+	+	+	-	+	-	+	-	-	+	+	-	+	+	-
13	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+
14	+	+	+	-	+	+	-	+	-	+	+	-	-	+	+
15	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+
16	+	+	+	-	+	-	+	-	-	+	+	-	+	+	-
17	+	+	+	+	-	+	-	-	+	+	-	+	-	+	-
18	+	+	+	-	+	+	-	-	-	+	+	-	-	+	-
19	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+
20	+	+	+	-	+	-	+	+	-	+	+	-	+	+	+
21	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+
22	+	+	+	-	+	+	-	-	-	+	+	-	-	+	-
23	+	+	+	+	-	-	+	-	+	+	-	+	+	+	-
24	+	+	+	-	+	-	+	-	-	+	+	-	+	+	-
25	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+

Задание 2. На основе данных из Российского статистического ежегодника. Раздел для исследования выбрать по согласованию с преподавателем. В платформе предусмотреть использовать скриптов на одном из языков аналитики, а также использование функций языка M, языка DAX, различные инструменты визуализации. В модели данных предусмотреть несколько таблиц и их взаимосвязь.

Типовые вопросы для опроса по теме 4:

1. Дайте общую характеристику средств бизнес-аналитики
2. Дайте характеристику моделей данных, используемых в Qlik
3. Дайте общую характеристику платформы Power Platform.
4. Выполните сравнительный анализ Power BI desktop, Power BI online
5. Дайте общую характеристику визуальных инструментов, которыеиспользуются п Power BI.
6. Какие графические приложения входят в состав Microsoft office. Дайте общуюхарактеристику языка DAX/

1. Использование матричных моделей в производственном планировании
2. Математическая модель межотраслевого (межпродуктового) баланса
3. Использование модели межпродуктового баланса в производственном планировании
4. Построение, расчет параметров и оптимизация сетевого графика комплекса работ
5. Классификация систем массового обслуживания и их показатели эффективности
6. Моделирование системы массового обслуживания: основные параметры, граф состояний
7. Вычисление вероятностей состояний системы массового обслуживания

Экзаменационные вопросы

Экзамен по дисциплине осуществляется в форме устного ответа и решения практической задачи на компьютере. Задание билета включает в себя два теоретических вопроса и одно практическое задание для решения в компьютерных программах визуализации и анализа данных. При проведении зачета с оценкой экзаменационный билет выбирает сам экзаменуемый в случайном порядке с помощью датчика случайных

чисел или с помощью выбора бланка билетов. Оценка результатов устного аттестационного испытания объявляется обучающимся в день его проведения. По окончании ответа на вопросы билета экзаменатор может задать экзаменуемому дополнительные и уточняющие вопросы в пределах учебного материала, вынесенного на экзамен. При оценке знаний магистранта преподаватель должен руководствоваться приведенной в рабочей программе шкалой оценивания результатов, а также следующими критериями:

– правильность ответа на вопросы билета;
– полнота ответа, знание терминологии;
– логика и аргументированность ответа;
– способность делать выводы.

Критерии и шкала оценивания результатов промежуточной аттестации (экзамен)

Обозначения	Формулировка требований к степени освоения

Цифра	Оценка	дисциплины
2	Неудовлетворительно	Оценка «неудовлетворительно» выставляется магистранту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями или неправильно решает задачу. Как правило, оценка «неудовлетворительно» ставится магистрантам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине. Учебные достижения в семестровый период и результаты рубежного контроля демонстрировали низкую степень овладения программным материалом.
3	Удовлетворительно	Оценка «удовлетворительно» выставляется магистранту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при решении практической задачи экзамена. Учебные достижения в семестровый период и результаты рубежного контроля демонстрируют достаточную (удовлетворительную) степень овладения программным материалом.
4	Хорошо	Оценка «хорошо» выставляется магистранту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положенияпри решении практических вопросов и задачи, вынесенной на экзамен, владеет необходимыминавыками и приемами их выполнения. Учебные достижения в семестровый период и результаты рубежного контроля демонстрируют хорошую степень овладения программным материалом учебной дисциплины.
5	Отлично	Оценка «отлично» выставляется магистранту,
		если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно излагает его на экзамене, умеет тесно увязывает теорию с практикой, свободно справляется с задачей, вынесенной на экзамен, правильно отвечает на дополнительные вопросы по теме билета. Учебные достижения в семестровый период и результаты рубежного контроля демонстрируют высокую степень овладения программным матер

Типовые вопросы, выносимые на экзамен

1. Понятие анализа данных. Стандартизация процесса анализа данных.
2. 2D и 3D-диаграммы. Параметры диаграмм. Построение диаграмм в Python и R.
3. Диаграммы рассеяния. Статистические диаграммы, гистограммы, функции ядерной плотности. Диаграммы размахов. Круговые и столбиковые диаграммы. Скрипичные диаграммы.
4. Предобработка данных. Очистка данных. Борьба с пропусками и аномалиями. Описательная статистика и подгонка распределений. Классические методы и критерии статистики.
5. Таблицы сопряженности. Номинальные, порядковые и интервальные статистики.
6. Методы повторной выборки. Бутстреп. Кросс-проверка. Перестановочный тест.Метод складного ножа.
7. Показатели графа. Размер. Плотность. Компоненты. Диаметр. Коэффициенткластеризации. Центральность.
8. Социальная сеть. Меры центральности. Центральность по близости.Центральность по посредничеству.
9. Кластеризация на графах. Построение графов в R. Библиотеки работы с графамии социальными сетями igraph, network.
10. Графическое представление и укладка сети. Алгоритмы укладки.
11. Социальная сплоченность. Клики. Обнаружение сообществ. Модулярность.
12. Сети аффилированности.
13. Динамические сети. Моделирование динамических сетей.
14. Агрегирование и группирование данных. Операторы агрегирования.
15. Сводная таблица. Power pivot. Power view.
16. Понятие проекции. Принцип ординации наблюдений. Методы ординации. Основные многомерные методы статистического анализа. Задачи снижения размерности данных.
17. Метод главных компонент. Факторный анализ. Основные этапы факторного анализа. Критерии факторного анализа. Матрица факторных нагрузок. Общностьи характерность переменных. График факторных нагрузок. Методы вращения
18. Многомерное шкалирование. Понятие стресса. Анализ соответствий.
19. Общая характеристика средств бизнес-аналитики.
20. Платформа Power BI. Основные задачи внедрения Power BI в компании. Составплатформы. Редактор Power BI querry Editor.
21. Язык DAX. Средства визуализации данных.
22. Платформа Qlik sense. Особенности построения отчетов с помощью Qlik sense.
23. Общая характеристика Tableau.

Типовые задачи на экзамен

Задача 1. Создать случайную последовательность в R размером в 500 наблюдений с использованием генератора равномерно распределенных чисел в диапазоне от 0 до 10. Проверить статистическую гипотезу о числовых значениях параметров:

1 $H_0 : a = 0{,}5; H_1 : a \neq 0{,}5$.

2 $H_0 : a = 5; H_1 : a > 5$.

Построить гистограмму распределения в R, SPSS, python. Построить гистограмму частот и гистограмму относительных частот. При построении гистограммы оценить и задать число интервалов. Указать название осей и название гистограммы, а также заливку синего цвета. На диаграмму поместить кривую ядерной плотности, а также аппроксимацию равномерным законом распределения. При построении кривой регулировать ее гладкость.

- Оценить статистические характеристики.
- При проверке гипотезы: использовать одновыборочный Т-критерий. Задать уровень значимости 0,05. Использовать одностороннюю и двухстороннюю проверки гипотезы.
- Проверить гипотезу о равномерном законе распределения с помощью критерияКолмогорова-Смирнова.

Задачу решить в различных приложених.

Задача 2. В файле ГосСлужба.txt приведены данные по стажу работы, стажу в должности и возрасту в государственной службе.

–Построить гистограммы распределения случайных величин.

–Оценить выборочные характеристики.

–Проверить статистические гипотезы о значимом отличии стажа в должности, стажа работы на гос. службе и возраста для мужчин и женщин с использованием t-критерия и критерия Манна-Уитни.

–Построить диаграммы размаха для случайных величин: возраст, стаж службы.Задачу решить в SPSS, python, R.

Задача 3. Таксомоторную компанию интересует зависимость между средним пробегом автомашины в расчете на 1 л топлива и возрастом машины. Были взяты 12 автомашин одной марки. Поскольку водителями были мужчины и женщины, предполагалось, что какая-то часть изменчивости пробега определяется разной техникой вождения у мужчин и женщин. Значения среднего пробега были рассчитаны на основе сведений о расходе горючего после прохождения машиной расстояния 100 км. Данные приведены в таблице.

Пол (мужчины, женщины)	Возраст машины, лет	Расход горючего, км.
мужчина	3	8,92
женщина	4	8,8
женщина	3	9,48
мужчина	2	9,68
женщина	1	10,2
мужчина	5	8,44
мужчина	4	8,24
мужчина	1	9,6
женщина	1	10,4
мужчина	2	9,24
женщина	2	9,92
мужчина	3	8,08

- Определить, значимы ли различия между пробегом для водителей-мужчин и водителей женщин, используя Т-тест для независимых групп (двухсторонний и односторонний). Для проверки гипотезы проверить гипотезу о постоянстве дисперсии.

Сравнить результаты проверки гипотезы с результатами проверки по критерию Манна-Уитни. Построить диаграммы размаха.

- Построить ящичные диаграммы для водителей мужчин и водителей-женщин.
- Решить задачу построения описательной статистики в различных приложениях.

Для проверки гипотезы по критерию Манна-Уитни в R использовать функцию wilcox.test($y \sim x$, data)

Задача 4. Создать две случайные последовательности двух случайных величинах, размером в 200 наблюдений, полученных с помощью генераторов нормально распределенных случайных чисел, имеющих одинаковое математическое ожидание, равное5 и ско, соответственно 1 и 2.

- Проверить гипотезу о равенстве математических ожиданий и дисперсий данныхвеличин.
- Изменить генератор, добавив в первый генератор смещение математическогоожидания. Вновь проверить статистическую гипотезу.
- Проверить гипотезы о нормальном законе распределения.
- Найти сумму пяти случайных величин, равномерно распределенных на интервале 0,

2. Проверить гипотезу о нормальном законе распределения суммы.

Задачу решить с помощью статистических критериев в R, SPSS, python. Построитьвероятностные и квантиль-квантиль графики.

Задача 5. Решить задачу анализа данных (Kwan K.C. et al., 1976) по скорости выведения из организма человека индометацина – одного из наиболее активных противовоспалительных препаратов. Данные находятся в наборе данных Indometh. The Indometh data frame has 66 rows and 3 columns of data on the pharmacokinetics of indometacin (or, older spelling, ‘indomethacin’).

Данный набор записать во внешний файл.

- Построить диаграмму зависимости концентрации в организме от времени вывода.
- Построить ящичную диаграмму для каждого испытуемого;
- Выявить аномальные наблюдения;
- Найти средние значения для каждой группы с помощью функции
- Построить гистограммы распределения.
- Построить ящичные диаграммы для каждого
- Найти описательные статистики.
- Проверить гипотезы о законах распределения.Задачу решить в различных приложениях.

Задача 6. В наборе данных InsectSprays (A data frame with 72 observations on 2 variables.), хранятся результаты эксперимента по изучению эффективности шести видов инсектицидных средств.

-Построить ящики с усами для каждого спрея.

-Построить описательную статистику.

-Проверить гипотезу о значимости отличий между результатами использования инстецидов A, C; A, F; E, D с помощью ящичной диаграммы.

При решении задачи использовать различные приложения. Для этого набор данных записать во внешний файл.

Задача 15. Решить задачу проверки статистической гипотезы о суточном расходе энергии у худощавых женщин и женщин с избыточным весом, если данные находятся в наборе energy (The energy data frame has 22 rows and 2 columns. It contains data on the energyexpenditure in groups of lean and obese women.) библиотеки ISwR. При проверке гипотезы использовать t-тест и тест Манна-Уитни. Построить ящичные диаграммы, гистограммы распределения. Выявить аномальные наблюдения. Проверить гипотезу о постоянстве дисперсии. Задачу решить в различных приложениях. Для этого набор данных записать вовнешний файл.

Задача 7. Решить задачу корреляционного анализа для данных из набора Boston

{MASS}

- Построить диаграмму рассеяния.

-Построить коррелограмму для корреляции с помощью функции corrgram и corrplot.

-Скрыть верхний треугольник коррелограммы. В нижнем треугольнике использоватьэллипсы.

-Найти частные коэффициенты корреляции.

-Найти корреляцию Пирсона, Спирмена, Кенделла.Задачу решить в различных приложениях.

Задача 8. Решить задачу факторного анализа для набора данных diamonds {ggplot2},

используя только числовые атрибуты. Задачу решить в SPSS, R, python, Orange. Прирешении задачи построить диаграмму каменистой осыпи.

Задача 9. Решить задачу кластерного анализа для набора данных state.x77. Задачу решить с помощью метода иерархической кластеризации, а также с помощью метода k- средних. Задачу решить в различных приложениях.

Задача 10. Решить задачу факторного анализа для набора данных States {carData}. Задачу решить в SPSS, R, python, Orange. При решении задачи построить диаграмму каменистой осыпи.

Описание системы оценивания

Оценочные средства (формы текущего и промежуточного контроля)	**Показатели оценки**	**Критерии оценки**
Опрос	Корректность и полнота ответов	**Сложный вопрос:** полный, развернутый, обоснованный ответ – 4 балла Правильный, но не аргументированный ответ – 2 балла Неверный ответ – 0 баллов **Обычный вопрос:** полный, развернутый, обоснованный ответ – 4 балла Правильный, но не аргументированный ответ – 2 балла Неверный ответ – 0 баллов. **Простой вопрос:** Правильный ответ – 2 балла; Неправильный ответ – 0 баллов
Тест	1) Правильность решений; 2) Корректность ответов	В зависимости от семестра максимальное количество баллов за один тес составляет 5 или 10 баллов
Задание	1) правильность решения; 2) корректность выводов 3) обоснованность решений	При условии 2 контрольных в семестре, максимальное количество баллов за каждую из них – 10. Если контрольная работа состоит из 5 заданий, то баллы за каждое из них начисляются от 0 до 2