

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Бианкина Алена Олеговна
Должность: Ректор
Дата подписания: 14.07.2023 14:54:31
Уникальный программный ключ:
b2aeadef209e4ec32d89f812db7eed614bb00b0c

Автономная некоммерческая организация высшего образования
« Институт социальных наук»



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.01 Анализ и визуализация данных

(краткое наименование дисциплины)

38.04.05 Бизнес-информатика

(код, наименование направления подготовки)

«Бизнес-аналитика»

(направленность)

магистр

(квалификация)

очная

(форма обучения)

Москва 2022

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы	4
2. Объем и место дисциплины в структуре образовательной программы.....	5
3. Содержание и структура дисциплины	5
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине	7
4.1. Формы и методы текущего контроля успеваемости обучающихся и промежуточной аттестации.....	7
4.2. Материалы текущего контроля успеваемости обучающихся.....	7
4.3. Оценочные средства для промежуточной аттестации.....	20
4.4. Методические материалы.....	26
5. Методические указания для обучающихся по освоению дисциплины	27
6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине.....	28
6.1. Основная литература.....	28
6.2. Дополнительная литература.....	29
6.3. Учебно-методическое обеспечение самостоятельной работы.....	29
6.4. Нормативные правовые документы.....	29
6.5. Интернет-ресурсы.....	29
6.6. Иные источники.....	29
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы	30

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

Дисциплина «Анализ и визуализация данных» обеспечивает овладение следующими компетенциями:

Таблица 1

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКС-2	Способен обосновывать подходы, используемые в бизнес-анализе, руководить и управлять бизнес-анализом с использованием информационно-коммуникационных технологий	ПКС-2.1	Способен использовать современные методы, информационные технологии, программный инструментарий в объеме, необходимом для решения задач бизнес-аналитики, использовать англоязычную документацию и справочные системы

В результате освоения дисциплины у магистрантов должны быть сформированы компетенции:

Таблица 2

ОТФ/ТФ (при наличии профстандарта)/ профессиональные действия	Код этапа освоения компетенции	Результаты обучения
Управление бизнес-анализом	ПКС-2.1	<p>на уровне знаний:</p> <p>Знать:</p> <ul style="list-style-type: none"> – методы анализа данных и машинного обучения; – возможности программных средств статистической обработки и визуализации данных <p>на уровне умения:</p> <p>Уметь:</p> <ul style="list-style-type: none"> – применять программные средства анализа данных, поддержки принятия решений; – применять современные средства бизнес-аналитики, разрабатывать интерактивные отчеты, dashboard, пользоваться ими при решении задач бизнес-анализа <p>Владеть:</p> <ul style="list-style-type: none"> - современными инфокоммуникационными технологиями, средствами и пакетами анализа данных и визуализации;

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

Общая трудоемкость дисциплины составляет 4 зачетные единицы /144 часа.

Таблица 3

Очная форма

Вид работы	Трудоемкость (акад/астр. часы)
Общая трудоемкость	144/108
Контактная работа с преподавателем	50/37,5
Лекции	20/15
Практические занятия	28/21
Консультация	2/1,5
Самостоятельная работа	58/43,5
Контроль	36/27
Формы текущего контроля	Задания, УО, Т
Форма промежуточной аттестации	Экзамен

Место дисциплины в структуре образовательной программы

Дисциплина «Анализ и визуализация данных» относится к части, формируемой участниками образовательных отношений образовательной программы подготовки магистранта учебного плана по направлению «Бизнес-информатика» 38.04.05. Преподавание дисциплины опирается на дисциплины программы бакалавриата «Эконометрика», «Анализ данных», «Теория вероятностей», «Математическая статистика», а также на дисциплины Б1.В.02 «Математические методы статистической обработки и анализа данных», Б1.О.07 «Аналитическая поддержка принятия решений», ФТД.02 «Предсказательная аналитика», Б1.В.09 «Интеллектуальный анализ данных, текстов и изображений».

В свою очередь она создаёт необходимые предпосылки для освоения программ таких дисциплин, как Б1.В.04 «Управление портфелем продуктов и ИТ-услуг».

Дисциплина закладывает теоретический и методологический фундамент для овладения умениям и навыками в ходе Б2.О.02 (Н) «Научно-исследовательская работа».

Знания, умения и навыки, полученные при изучении дисциплины, используются студентами при выполнении выпускных квалификационных работ.

3. Содержание и структура дисциплины

№ п/п	Наименование тем	Объем дисциплины, час.					Форма текущего контроля успеваемость и**, промежуточ ной аттестации* **	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий			СР		
			Л	ПЗ	КСР	СРО		СП
Тема 1	Анализ данных. Построение графиков и визуализация	26	6	8		12		УО/Зад/Т
Тема 2	Методы сетевого анализа	28	4	8		16		УО/Зад
Тема 3	Агрегирование и проецирование	28	4	8		16		УО/Зад/Т

	данных							
Тема 4	Средства бизнес-аналитики	26	4	8		14		УО/Зад
Промежуточная аттестация		36/27			2	36		Экзамен
Всего (акад./астр. часы):		144/108	20/15	28/21	2/1,5	58/43,5	36/27	

Примечание:

Используемые сокращения:

Л – занятия лекционного типа (лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях, обучающимся) ;

ПЗ – практические занятия (виды занятия семинарского типа за исключением лабораторных работ) ;

КСР – индивидуальная работа обучающихся с педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях (в том числе индивидуальные консультации) ;

СР – самостоятельная работа, осуществляемая без участия педагогических работников организации и (или) лиц, привлекаемых организацией к реализации образовательных программ на иных условиях;

СП – самопроверка;

СРО – самостоятельная работа обучающегося

контрольные работы (К), опрос (О), тестирование (Т)

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 query 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.

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине

4.1. Формы и методы текущего контроля успеваемости обучающихся и промежуточной аттестации.

В ходе реализации дисциплины используются следующие методы текущего контроля успеваемости обучающихся:

Таблица 4.1

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

4.1.2 Экзамен проводится по итогам текущего контроля

При выставлении оценок на экзамене учитывается активность магистрантов в течение занятий, выполнение задач, решаемых в ходе занятия, а также

4. 2. Материалы текущего контроля успеваемости обучающихся.

Типовые оценочные материалы по теме 1

Задания по теме 1

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

Задание 2. Воспользоваться графическими приложениями RCmdr, Rattle для решения задач графического анализа

Задание 3. Исследовать возможности графического пакета matplotlib, seaborn, Для исследования 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 последовательность из 100 чисел, заключенную от 0 до 10. Используя оператор цикла for и итератор i сформировать последовательность из 100 чисел от -1 до 9. Построить точечные, линейные диаграммы 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','Book Value'],  
                 index_col='Symbol')
```

```
sp500.head()
```

Провести анализ файла.

С помощью метода iloc отобразить только 10,20,30,40,50, 60 и 70 элементы

Отобразить 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 методом 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, полученными вручную.

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 использования диаграмм для решения различных задач графического анализа.
4. Какие задачи решаются во время предобработки данных. Назовите статистические гипотезы, которые могут проверяться при решении задач разведочного анализа.
5. Дайте общую характеристику программных платформ работы с графикой.
6. Дайте характеристику графической грамматики. Приведите примеры ее использования для построения различных диаграмм
7. Назовите статистические диаграммы. Что представляет собой диаграмма Кливленда?
8. Что такое повторная выборка? Какие методы применяются при решении задач повторной выборки?

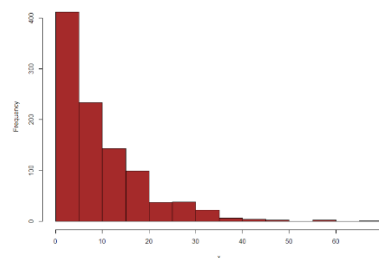
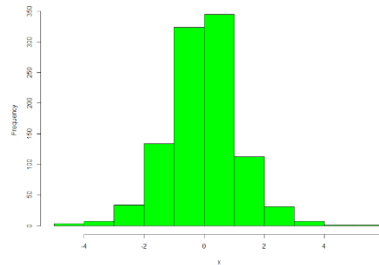
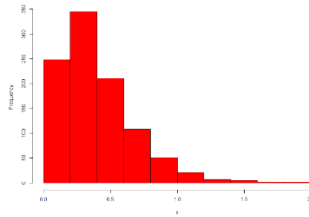
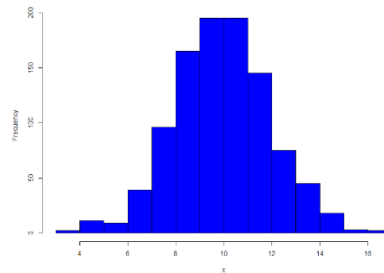
Тест

1. Гистограммы распределения

Поставить соответствие между заданными последовательностями и гистограммами распределения, если последовательности представляют собой случайные числа, полученные с помощью 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 позволяет проверить гипотезу о нормальном законе распределения

- Колмогорова-Смирнова
- Пирсона (Chi 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 для двух случайных величин/ Ответ дать с точностью до двух знаков после запятой

Таблица

наблюдений

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 пропуски заменены 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 величин

Таблиц

а
наблюдений

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

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 дисперсии

Данные заданы двумя последовательностями

$x \leftarrow c(0,2,3,1,4,5)$

$y \leftarrow c(3,4,5,3,4,3)$

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

$x \leftarrow c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)$

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

$x \leftarrow c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)$

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

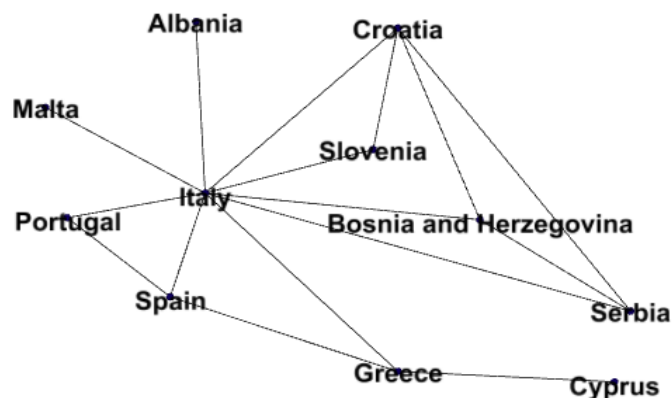
Типовые оценочные материалы по теме 2

Задания по теме 2

Задание 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: Анализ 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 3

Задания 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 различия 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 менеджер	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

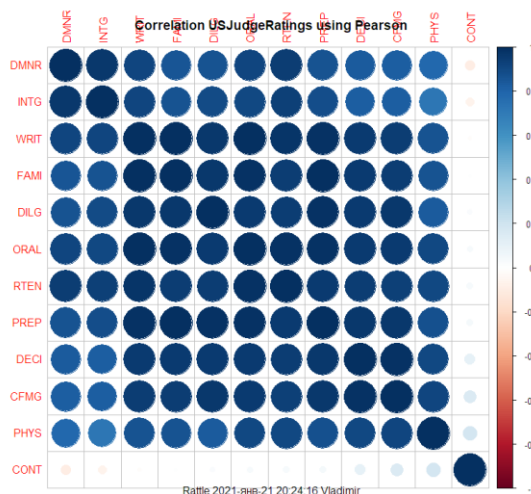
1. Этапы факторного анализа

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 выбрать подмножество анализируемых переменных
3. Целесообразно решить задачу факторного анализа

4. Для принятия решения о целесообразности факторного анализа необходимо использовать непараметрические критерии корреляции
5. Факторный анализ не целесообразен. Нужно попробовать решить задачу методом главных компонент

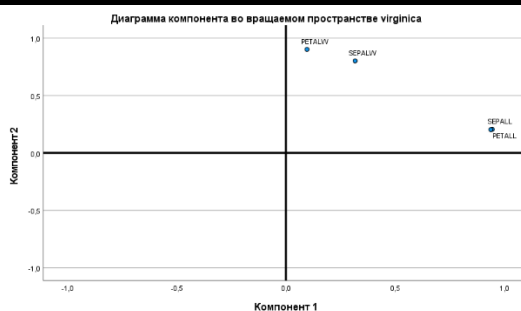
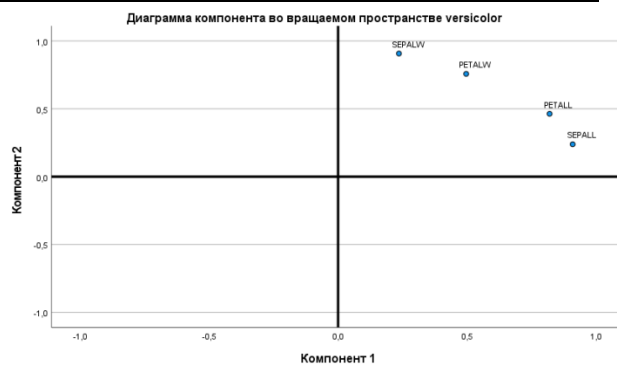
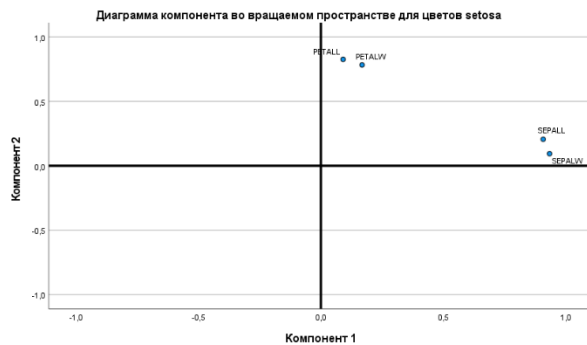
3. Матрица факторных нагрузок

Что означает элемент матрицы факторных нагрузок a_{ij} ?

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. стандартизованное значение переменной

4. Оценка качества факторного анализа

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



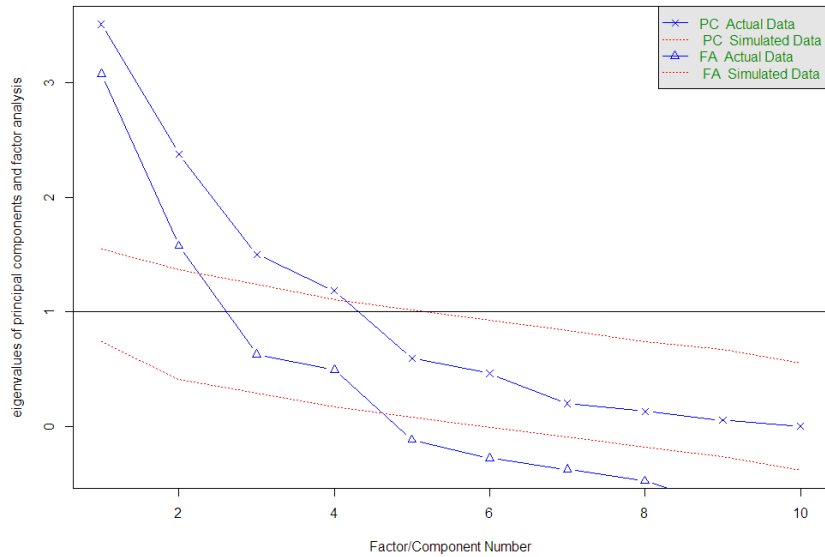
Выберите правильное ранжирование качества решения задачи факторного анализа для разных классов цветов

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

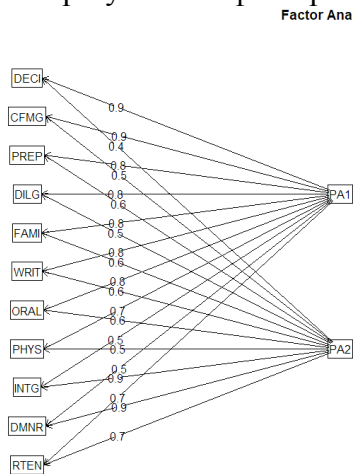
Диаграмма каменной осыпи



Чему равно рекомендуемое число главных компонент для фактических данных, если используется 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 первое число пары 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

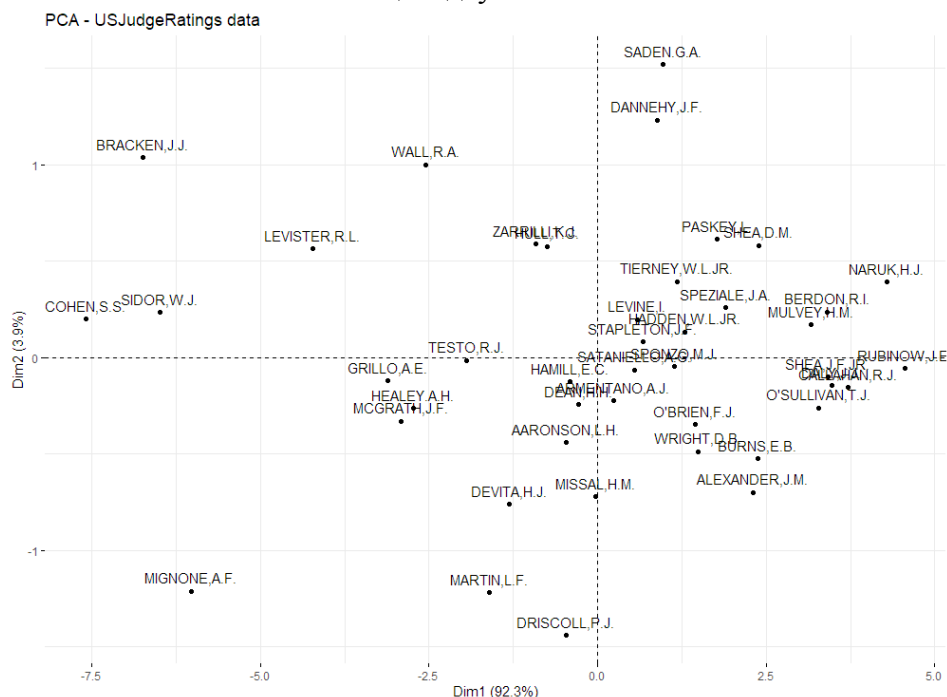
Для решения задачи факторного анализа в R выполнен следующий скрипт

```
library(psych)
```

`pc <- principal(USJudgeRatings[, -1], nfactors=6, rotate='none')`. Сколько главных компонент 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

Компонент	Начальные собственные значения			Извлечение суммы квадратов нагрузок		
	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %
1	2,911	72,770	72,770	2,911	72,770	72,770
2	,921	23,031	95,801	,921	23,031	95,801
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 с точностью 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	г	область	регион	городское население	зарплата	миграция	Млад. Смертность	Рождаемость	Смертность	население	ос. Фонды	преступления	сельское хозяйство	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, которые используются в Power BI.
6. Какие графические приложения входят в состав Microsoft office. Дайте общую характеристику языка DAX/

4.3. Оценочные средства для промежуточной аттестации.

Таблица 4.2

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКС-2	Способен обосновывать подходы, используемые в бизнес-анализе, руководить и управлять бизнес-анализом с использованием информационно-коммуникационных технологий	ПКС-2.1	Способен использовать современные методы, информационные технологии, программный инструментарий в объеме, необходимом для решения задач бизнес-аналитики, использовать англоязычную документацию и справочные системы

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 дисциплины
Цифра	Оценка	
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 положения при решении практических вопросов и задачи, вынесенной на экзамен, владеет необходимыми навыками 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
--	--	--

Показатели и критерии оценивания компетенций на различных этапах их формирования

Таблица 4.3

Код компонента компетенции	Показатель оценивания	Критерий оценивания
ПКС-2.1	Способен использовать современные методы, информационные технологии, программный инструментарий в объеме, необходимом для решения задач бизнес-аналитики, использовать англоязычную документацию и справочные системы	Использует информацию, методы и программные средства ее сбора, обработки и анализа, в том числе с использованием интеллектуальных методов. Самостоятельно решает задачи интеллектуального анализа данных, текстов и изображений с помощью аналитических платформ, фреймворков Python Keras, TensorFlow

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. Коэффициент кластеризации. Центральность.
8. Социальная сеть. Меры центральности. Центральность по близости. Центральность по посредничеству.
9. Кластеризация на графах. Построение графов в R. Библиотеки работы с графами и 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. Общность и 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. Состав платформы. Редактор Power BI query 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 T-критерий. Задать уровень значимости 0,05. Использовать одностороннюю и двухстороннюю проверки гипотезы.

- Проверить гипотезу о равномерном законе распределения с помощью критерия Колмогорова-Смирнова.

Задачу решить в различных приложениях.

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 T-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 ~ 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, равное 5 и ско, соответственно 1 и 2.

- Проверить гипотезу о равенстве математических ожиданий и дисперсий данных величин.

- Изменить генератор, добавив в первый генератор смещение математического ожидания. 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. Построить вероятностные 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 ящички с усами для каждого спрея.

- Построить описательную статистику.

- Проверить гипотезу о значимости отличий между результатами использования инсектицидов 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 energy expenditure 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 во внешний 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 использовать эллипсы.

- 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. При решении 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	<p>Сложный вопрос: полный, развернутый, обоснованный ответ – 4 балла Правильный, но не аргументированный ответ – 2 балла Неверный ответ – 0 баллов</p> <p>Обычный вопрос: полный, развернутый, обоснованный ответ – 4 балла Правильный, но не аргументированный ответ – 2 балла Неверный ответ – 0 баллов.</p> <p>Простой вопрос: Правильный ответ – 2 балла; Неправильный ответ – 0 баллов</p>
Тест	1) Правильность решений; 2) Корректность ответов	В зависимости от семестра максимальное количество баллов за один тест 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

4.4. Методические материалы по освоению дисциплины

Рабочей программой дисциплины предусмотрены следующие виды аудиторных занятий: лекции, практические занятия. На лекциях рассматриваются наиболее сложный материал дисциплины. Для развития у магистрантов креативного мышления и логики в каждой теме учебной дисциплины предусмотрены теоретические положения, инструментальные средства, а также примеры их использования при решении задач предиктивной аналитики. Кроме того, часть теоретического материала предоставляется на

самостоятельное изучение по рекомендованным источникам для формирования навыка самообучения.

Практические занятия предназначены для самостоятельной работы магистрантов по решению конкретных задач. Каждое практическое занятие сопровождается заданиями, выдаваемыми магистрантам для решения во внеаудиторное время.

Для работы с печатными и электронными ресурсами имеется возможность доступа к электронным ресурсам. Организация работы магистрантов с электронной библиотекой указана на сайте института).

5. Методические указания для обучающихся по освоению дисциплины

Обучение по дисциплине «Анализ и визуализация данных» предполагает изучение курса на аудиторных занятиях (лекции, практические работы) и самостоятельной работы обучающихся. Семинарские занятия дисциплины «Анализ и визуализация данных» предполагают их проведение в различных формах с целью выявления полученных знаний, умений, навыков и компетенций с проведением контрольных мероприятий. С целью обеспечения успешного обучения обучающийся должен готовиться к лекции, поскольку она является важнейшей формой организации учебного процесса, поскольку:

- знакомит с новым учебным материалом;
- разъясняет учебные элементы, трудные для понимания;
- систематизирует учебный материал;
- ориентирует в учебном процессе.

Подготовка к лекции заключается в следующем:

- внимательно прочитайте материал предыдущей лекции;
- узнайте тему предстоящей лекции (по тематическому плану, по информации лектора);
- ознакомьтесь с учебным материалом по рекомендуемой литературе;
- постарайтесь уяснить место изучаемой темы в своей профессиональной подготовке;
- запишите возможные вопросы, которые вы зададите лектору на лекции.

Подготовка к практическим занятиям:

- внимательно прочитайте материал лекций, относящихся к данному семинарскому занятию, ознакомьтесь с учебным материалом;
- ответьте на контрольные вопросы по семинарским занятиям, готовьтесь дать развернутый ответ на каждый из вопросов;
- уясните, какие учебные элементы остались для вас неясными и постарайтесь получить на них ответ заранее (до семинарского занятия) во время текущих консультаций преподавателя;
- готовиться можно индивидуально, парами или в составе малой группы, последние являются эффективными формами работы;
- рабочая программа дисциплины в части целей, перечню знаний, умений, терминов и учебных вопросов может быть использована вами в качестве ориентира в организации обучения.

Выполнение задания:

- выберите набор данных (временной ряд, временные ряды) для выполнения задания;
- выполните анализ используемых признаков (целевого признака);
- проанализируйте качество исходных данных;
- выполните выбор инструментов предобработки для улучшения качества исходных данных, а также формулировки предварительных гипотез;
- решите задачу прогнозирования уровней временного ряда;

- исследуйте возможность извлечения признаков временного ряда;
- решите задачу анализа выявленных признаков;
- оформите отчет по результатам выполнения задания.

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. 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: 12.11.2020). - Режим доступа: для авторизир. Пользователей.
7. Курносоев 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
8. Мастицкий С. Э. (2020) Анализ временных рядов с помощью R. — Электронная книга, адрес доступа: <https://ranalytics.github.io/tsa-with-r>
9. 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: 01.10.2020)
10. Мхитарян В. С., Архипова М. Ю., Дуброва Т. А., Миронкина Ю. Н., Сиротин В. П. Анализ данных. – М.: Юрайт, 2020 – 490 с. Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166> (дата обращения: 29.09.2020)
11. О'Нил, Кэти. Data Science : Инсайдерская информация для новичков. Включая язык R : [пер. с англ.] – СПб. Питер. – 368 с. Текст: электронный. - URL: <http://new.ibooks.ru/bookshelf/359209/reading> (дата обращения: 25.01.2021)
12. Сузи Р.А. Python. – СПб.: БХВ – Петербург, 2015 – 759 с. электронный ресурс Джеймс Г., Уиттон Д., Хастис 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>

Все источники основной литературы взаимозаменяемы.

6.2 Дополнительная литература

1. Гусарова Н.Ф, Анализ социальных сетей. Основные понятия и метрики. – СПб.: Университет ИТМО, 2016 – 67 с.
2. 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.
3. Ланц Б. – Машинное обучение на R/ - СПб. :Питер. – 2020 – 464 с.
4. 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.
5. Маккинни У. Python и анализ данных. – М.: ДМК, 2015 – 482 с.
6. Нильсен Эйлин. Практический анализ временных рядов: прогнозирование со статистикой и машинное обучение. –М.: ООО Диалектика – 2021 – 544 с.
7. 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.

8. 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.

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

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

6.3. Учебно-методическое обеспечение самостоятельной работы

Приказ Минобрнауки России от 19.11.2013 N 1259 (ред. от 05.04.2016) "Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам подготовки научно-педагогических кадров в аспирантуре (адъюнктуре)" (Зарегистрировано в Минюсте России 28.01.2014 N 31137)

6.4. Нормативные правовые документы

Не используются

6.5. 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 стандарты качества.
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 / Casewise Corporate 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 Business Process Analysis Suite
22. <http://www.softwareag.com/ru> / Software AG
23. <http://www-03.ibm.com/software/products/ru/ru/wbi> / IBM WebSphere Business Modeler

6.6. Иные источники.

Не используются.

7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Учебная дисциплина включает использование программного обеспечения Microsoft Excel, Microsoft Word, для подготовки текстового и табличного материала.

Интернет-сервисы и электронные ресурсы (поисковые системы, электронная почта, профессиональные тематические чаты и форумы, системы аудио и видео конференций, онлайн энциклопедии, справочники, библиотеки, электронные учебные и учебно-методические материалы).

Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

№ п/п	Наименование
1.	Компьютерные классы с персональными ЭВМ, объединенными в локальные сети с выходом в Интернет
2.	Пакет Excel -2016, professional plus, IBM SPSS statistics, R, RStudio, Anaconda
3.	Мультимедийные средства в каждом компьютерном классе и в лекционной аудитории
4.	Браузер, сетевые коммуникационные средства для выхода в Интернет. Сервисы и службы Azure

Компьютерные классы из расчета 1 ПЭВМ для одного обучаемого. Каждому обучающемуся должна быть предоставлена возможность доступа к сетям типа Интернет в течение не менее 20% времени, отведенного на самостоятельную подготовку.

