МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
имени И.Т. ТРУБИЛИНА»

Факультет прикладной информатики
Системного анализа и обработки информации

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ) «ОСНОВЫ АНАЛИЗА ДАННЫХ»

Уровень высшего образования: бакалавриат

Направление подготовки: 38.03.05 Бизнес-информатика

Направленность (профиль)подготовки: Анализ, моделирование и формирование интегрального представления стратегий и целей, бизнес-процессов и информационно-логической инфраструктуры предпри

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Год набора: 2024

Срок получения образования: 4 года

Объем: в зачетных единицах: 4 з.е.
в академических часах: 144 ак.ч.

2024

Разработчики:
Доцент, кафедра системного анализа и обработки информации Павлов Д.А.

Рабочая программа дисциплины (модуля) составлена в соответствии с требованиями ФГОС ВО по направлени ю подготовки 38.03.05 Бизнес-информатика, утвержденного приказом Минобрнауки от 29.07.2020 № 838, с учетом трудовых функций профессиональных стандартов: "Менеджер по информационным технологиям", утвержден приказом Минтруда России от 30.08.2021 № 588н; "Специалист по информационным системам", утвержден приказом Минтруда России от 13.07.2023 № 586н; "Системный аналитик", утвержден приказом Минтруда России от 27.04.2023 № 367н.

Согласование и утверждение

№	Подразделение или коллегиальный орган	Ответственное лицо	ФИО	Виза	Дата, протокол (при наличии)
1	Системного анализа и обработки информации	Заведующий кафедрой, руководитель подразделения, реализующего ОП	Барановская Т.П.	Согласовано	08.04.2024, № 8

1. Цель и задачи освоения дисциплины (модуля)

Цель освоения дисциплины - является изучение теоретических основ и методов анализа данных, применяемых при решении прикладных задач и использования их в современных информационных системах.

Задачи изучения дисциплины:

- изучение методов статистического анализа данных;
- изучение существующих технологий подготовки данных к исследованию и овладение практическими умениями и навыками реализации технологий анализа данных;
- формирование и проверка гипотез о природе и структуре данных.

2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы

Компетенции, индикаторы и результаты обучения

ОПК-4 Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений

ОПК-4.1 Понимает роль информации в процессе принятия управленческих решений и проводит оценку ее свойств

Знать:

ОПК-4.1/Зн1 Знает свойства информации

Уметь:

ОПК-4.1/Ум1 Умеет определить роль информации в процессе принятия управленческих решений и проводит оценку ее свойств

Владеть:

ОПК-4.1/Нв1 Владеет методами оценки информации в процессе принятия управленческих решений

ОПК-4.2 Применяет современные программные средства и методы сбора, обработки и анализа информации

Знать:

ОПК-4.2/Зн1 Знает современные программные средства и методы сбора, обработки и анализа информации

Уметь:

ОПК-4.2/Ум1 Умеет применять современные программные средства и методы для сбора, обработки и анализа информации

Владеть:

ОПК-4.2/Нв1 Владеет знаниями о современных программных средствах и методах сбора, обработки и анализа информации

ОПК-4.3 Использует экономико-математические модели и методы как средство информационно-аналитической поддержки принятия управленческих решений

Знать:

ОПК-4.3/Зн1 Знает методы экономико-математического моделирования

Уметь:

ОПК-4.3/Ум1 Умеет применять методы экономико-математического моделирования для информационно-аналитической поддержки принятия управленческих решений

Владеть:

ОПК-4.3/Нв1 Использует экономико-математические модели и методы как средство информационно-аналитической поддержки принятия управленческих решений

ОПК-4.4 Демонстрирует возможность программной реализации экономико-математических методов и моделей в системах поддержки принятия управленческих решений

Знать:

ОПК-4.4/Зн1 Знает возможности программной реализации экономико-математических методов и моделей в системах поддержки принятия управленческих решений

Уметь:

ОПК-4.4/Ум1 Умеет проводить программную реализацию экономико-математических методов и моделей в системах поддержки принятия управленческих решений

Владеть:

ОПК-4.4/Нв1 Демонстрирует возможность программной реализации экономико-математических методов и моделей в системах поддержки принятия управленческих решений

3. Место дисциплины в структуре ОП

Дисциплина (модуль) «Основы анализа данных» относится к обязательной части образовательной программы и изучается в семестре(ах): 5.

В процессе изучения дисциплины студент готовится к решению типов задач профессиональной деятельности, предусмотренных ФГОС ВО и образовательной программой.

4. Объем дисциплины и виды учебной работы

Период обучения	Общая трудоемкость (часы)	Общая трудоемкость (ЗЕТ)	Контактная работа (часы, всего)	Внеаудиторная контактная работа (часы)	Лабораторные занятия (часы)	Лекционные занятия (часы)	Самостоятельная работа (часы)	Промежуточная аттестация (часы)
Пятый семестр	144	4	69	3	32	34	48	Экзамен (27)
Всего	144	4	69	3	32	34	48	27

5. Содержание дисциплины

5.1. Разделы, темы дисциплины и виды занятий

(часы промежуточной аттестации не указываются)

Наименование раздела, темы		ная контактная абота	ные занятия	е занятия	льная работа	ые результаты оотнесенные с ии освоения

	Всего	Внеаудиторная	Лабораторные	Лекционные	Самостоятельная	Планируемые результаты обучения, соотнесенные с результатами программы
Раздел 1. Статистический анализ данных	**50**		**18**	**10**	**22**	ОПК-4.1 ОПК-4.2 ОПК-4.3
Тема 1.1. Введение в анализ данных	6		4	2		
Тема 1.2. Фреймворки Python: Numpy, Pandas	32		12	4	16	
Тема 1.3. Случайные величины и их характеристики	12		2	4	6	
Раздел 2. Разведочный анализ данных	**22**	**2**	**4**	**4**	**12**	ОПК-4.1 ОПК-4.2 ОПК-4.3 ОПК-4.4
Тема 2.1. Инструменты и методы разведочного анализа данных	22	2	4	4	12	
Раздел 3. Проверка статистических гипотез. Принятие решений	**45**	**1**	**10**	**20**	**14**	ОПК-4.1 ОПК-4.2 ОПК-4.3 ОПК-4.4
Тема 3.1. Выборочный метод	10		2	6	2	
Тема 3.2. Доверительные интервалы	14		4	6	4	
Тема 3.3. Проверка гипотез	14		4	6	4	
Тема 3.4. A/B тесты	7	1		2	4	
Итого	**117**	**3**	**32**	**34**	**48**	

5.2. Содержание разделов, тем дисциплин

Раздел 1. Статистический анализ данных
(Лабораторные занятия - 18ч.; Лекционные занятия - 10ч.; Самостоятельная работа - 22ч.)

Тема 1.1. Введение в анализ данных
(Лабораторные занятия - 4ч.; Лекционные занятия - 2ч.)
Понятие данных
Источники данных
Специальности в области науки о данных
Большие данные
Методы анализа данных
Дата-ориентированных подход в принятии решений

Тема 1.2. Фреймворки Python: Numpy, Pandas
(Лабораторные занятия - 12ч.; Лекционные занятия - 4ч.; Самостоятельная работа - 16ч.)
Numpy
Pandas
Визуализация данных
Агрегация и группировка данных

Тема 1.3. Случайные величины и их характеристики
(Лабораторные занятия - 2ч.; Лекционные занятия - 4ч.; Самостоятельная работа - 6ч.)

Описательные статистики
Гистограмма и эмпирическая функция распределения
Распределения и описательные статистики
Зависимые и независимые случайные величины
Ковариация и корреляция
Нормальное распределение и его свойства
Центрирование и нормирование

Раздел 2. Разведочный анализ данных
(Внеаудиторная контактная работа - 2ч.; Лабораторные занятия - 4ч.; Лекционные занятия - 4ч.; Самостоятельная работа - 12ч.)

Тема 2.1. Инструменты и методы разведочного анализа данных
(Внеаудиторная контактная работа - 2ч.; Лабораторные занятия - 4ч.; Лекционные занятия - 4ч.; Самостоятельная работа - 12ч.)
Предобработка данных
Выявление аномалий и выбросов
Идентификация связей и корреляций между переменными
Преобразование данных
Визуализация

Раздел 3. Проверка статистических гипотез. Принятие решений
(Внеаудиторная контактная работа - 1ч.; Лабораторные занятия - 10ч.; Лекционные занятия - 20ч.; Самостоятельная работа - 14ч.)

Тема 3.1. Выборочный метод
(Лабораторные занятия - 2ч.; Лекционные занятия - 6ч.; Самостоятельная работа - 2ч.)
Схема математической статистики
Мощь средних
Разность средних
Мощь долей
Число наблюдений

Тема 3.2. Доверительные интервалы
(Лабораторные занятия - 4ч.; Лекционные занятия - 6ч.; Самостоятельная работа - 4ч.)
Асимптотические доверительные интервалы
Точные доверительные интервалы

Тема 3.3. Проверка гипотез
(Лабораторные занятия - 4ч.; Лекционные занятия - 6ч.; Самостоятельная работа - 4ч.)
p - value
Виды критериев
Гипотезы о долях
Гипотезы о средних
Гипотезы о дисперсии

Тема 3.4. A/B тесты
(Внеаудиторная контактная работа - 1ч.; Лекционные занятия - 2ч.; Самостоятельная работа - 4ч.)
Схема сплит тестирования
Проблемы тестов

6. Оценочные материалы текущего контроля

Раздел 1. Статистический анализ данных

Форма контроля/оценочное средство: Расчетно-графическая работа

Вопросы/Задания:

1. Выполните задания в Jupyter Notebook

Задания

Выполните задания

1. Числовые и строковые типы данных языка Python. Управляющие конструкции.

1.1. Работа с числами. Базовые числовые типы int и float

1. Значения переменных A и B ввести с клавиатуры и вывести на экран. После этого значения меняются местами, т.е. A нужно присвоить значение B, а B – значение A, и вновь значения переменных вывести на экран.

2. Значение x вводится с клавиатуры. Вычислите $y=(x+1)^2\cdot \sqrt{x^3+1}$ Выведите на экран значения x и y с тремя знаками после запятой.

3. Вычислите сумму цифр пятизначного числа.

4. Для заданного трехзначного числа выведите число, у которого цифры идут в обратном порядке, например, для числа 123 ответ 321.

5. Ввести координаты 2 точек: (x_1,y_1) и (x_2,y_2). Вычислите расстояние между этими точками. Результат выведите с 5 знаками после запятой.

6. Разработать программу вычисления по известному радиусу площади круга и длины окружности.

7. Разработать программу по вычислению площади кольца по известным значениям его внешнего и внутреннего радиусов.

8. Разработать программу вычисления радиуса круга и его площади по известной длине окружности.

9. Разработать программу вычисления объема шара: а) по радиусу; б) по длине его экваториальной параллели. Объём шара рассчитывается по формуле $V=\frac{4}{3}\pi{r^3}$, где r - радиус.

10. Имеются два n–мерных вектора x и y, которые задают координаты n точек на плоскости (случайные целые числа). Найти наиболее близкие друг другу точки.

11. Треугольник задан координатами своих вершин: $(x_1,y_1),(x_2,y_2),(x_3,y_3)$. Значения координат определите с помощью присваивания. Они могут быть нецелыми. Найти периметр и площадь треугольника.

12. Пусть пользователь вводит 5 чисел: a, b, c, d, e. Реализуйте программу расчета выражения вида $res=\frac{a+bc}{2d-e}+Ost[b^e/c]$, где Ost - остаток от деления. Учесть невозможность деления на 0.

1.2. Строки, функции и методы работы со строками

1. Введите строку, состоящую из двух цифр. Преобразуйте ее в целое и вещественное число. Выведите полученные три значения (строку, целое число, вещественное число) на экран в одной строке через запятую, затем пропустите строку и вновь выведите значения по одному на строке. Перед каждым значением выведите его тип.

2. Для строки 'Финансовый университет' двумя способами получить подстроку с 1-го по 4-й символы включительно.

3. Составить строку из всех четных символов строки 'Финансовый университет'.

4. Получить подстроку неизвестной заранее строки, содержащую половину символов строки и расположенную посередине строки.

5. Коротко записать создание строки 'oneoneoneoneoneonetwotwotwo'

6. Для введенной строки выведите (на отдельной строке):
- второй символ этой строки;
- предпоследний символ этой строки;
- всю строку, кроме последних двух символов;
- все символы с четными индексами (считая, что индексация начинается с 0, поэтому символы выводятся, начиная с первого);
- все символы с нечетными индексами, то есть начиная со второго символа строки;
- все символы в обратном порядке;
- все символы строки через один в обратном порядке, начиная с последнего;
- длину данной строки.

7. Не используя метод count, для заданной строки выполните:
- если символ * в данной строке отсутствует, выведите текст «нет символа»;
- если символ * встречается в строке только один раз, выведите его индекс;
- если символ * встречается два и более раз, выведите индекс его первого, второго и последнего вхождения, удалите первый и последний символ * из строки.

8. Дана строка, состоящая из слов, разделенных пробелами. В этой строке:
- удалите все лишние пробелы (в начале, в конце, между словами оставить ровно один пробел);
- поменяйте регистр символов (строчные сделать прописными, прописные – строчными);
- определите, сколько в ней слов.

9. Строка содержит фамилию, имя и отчество, записанные через пробелы. Например «Иванов Иван Иванович». Для этой строки:
- выведите информацию в виде:
```
Фамилия       Иванов
Имя           Иван
Отчество      Иванович
```
- получите строки вида «Иванов И.И.» и «И.И. Иванов».

10. Напишите программу, которая осуществляет вывод кошки на экран. Примечание: кошка выглядят примерно так:
```
  /\_/\
  /@ @\
 (~~0~~)
```

```

11. Для двух произвольных строк провести сравнение на совпадение содержимого без учета регистра букв и начальных и конечных пробельных символов. Например, для двух строк ' HeLLO WOrlD' и 'hello WORLD ' такое сравнение должно возвращать True (содержание строк совпадает).

12. По номеру года определите, является ли данный год високосным. (год является високосным, если его номер кратен 4, но не кратен 100, а также если он кратен 400).

18. Даны точки $A=(1,1),B=(5,5),C=(1,4),D=(7,2)$. Найти точку пересечения отрезков $AB$ и $CD$.

2. Выполните задания в Jupyter Notebook

Выполните задания

1.3. Операторы сравнения. Логические операторы. Инструкция ветвления if...else
Определить время года по номеру месяца.
Определить минимальное значение среди чисел
(не использовать стандартные функции max и min).
Напишите программу, которая переводит оценку из 100-балльной системы в пятибалльную по правилам, принятым в университете.
Вычислить значение
, если цифра
входит в запись заданного трехзначного числа, и
– в противном случае.
Даны два отрезка
и
. Найдите их пересечение. Если отрезки не пересекаются, то выдайте сообщение.
По номеру года определите, является ли данный год високосным. (год является високосным, если его номер кратен 4, но не кратен 100, а также если он кратен 400).
Для отрезков длины a, b, c определить, можно ли из них составить треугольник и является ли этот треугольник прямоугольным.
Напишите программу для решения уравнения
Коэффициенты
могут быть любыми числами.
Вводится целое число. Выведите его на экран и допишите к нему слова «рубль», «рубля» или «рублей» в зависимости от значения. Алгоритм:
исключение: если число оканчивается на 11, 12, 13 или 14, добавляем слово «рублей»;
если число оканчивается на 1, добавляем слово «рубль»;
если число оканчивается на 2, 3 или 4, добавляем слово «рубля»;
если число оканчивается на цифры 5, 6, 7, 8, 9 или 0, добавляем слово «рублей».
Пользователь поочередно вводит координаты точки в декартовой системе координат. Определить, какой четверти принадлежит данная точка или на какой оси она находится. Расположение точки вывести на экран. Найти произведение номера четверти на расстояние от этой точки до начала координат и вывести его на экран. Если точка лежит на оси, считать, что номер четверти равен 0.
Выведите значение заданного целого числа от 0 до 999 прописью. Например, «сто девяносто один» для числа 191, «одиннадцать» для числа 11.
Вычислите значение выражения, которое состоит из целых чисел и знаков «+» и «-». Выражение вводится как символьная строка.
Пользователь поочередно вводит 2 строки. Определите, какая строка длиннее и на сколько символов, а какая строка стоит раньше в лексикографическом порядке. Собрать результирующую строку, которая бы содержала 1 и 2 строки, разделенные переносом строки.
Реализовать калькулятор, который принимает от пользователя через пробел строку следующего вида: «a op b», где a и b – некоторые числа, а ‘op’ определяет оператор и может
```

принимать значения «
. В зависимости от оператора с помощью форматирования строк вывести результат в виде: «a + b = 3 + 2 = 5», где была получена строка «3 + 2». Для операций возведения в степень и деления по модулю использовать вместо знака оператора соответствующие выражения. Например, для строки «2 ** 3» должно быть выведено «a в степени b = 2 в степени 3 = 8».
Напишите программу, которая в зависимости от введенного пользователем числа N, осуществляет вывод N кошек на экран. Пусть 1<= N <=10. Примечание: кошки при N=3 выглядят примерно так

```
 /\__/\ /\__/\ /\__/\
 /@ @\ /@ @\ /@ @\
 (~~0~~) (~~0~~) (~~0~~)
```

Имеются два n – мерных вектора x и y, которые задают координаты n точек на плоскости (случайные целые числа). Найти наиболее близкие друг другу точки.
Из множества целых чисел от 1 до N выделить множество N2 числа, кратные 2, множество N3 – кратные 3, множество N 6 – кратные 6 (т.е. кратные и 2 и 3), множество N 23 – кратные либо 2, либо 3.
Даны m (m>1) слов. Найти общее количество заданной буквы в этих словах.
Даны m (m>1) слов. В каком из них доля (в %) заданной буквы больше.

3. Выполните задания в Jupyter Notebook

Выполните задания

Имеется список из 20 случайных чисел от 0 до 100. Напишите функцию, которая разбивает этот список на
списков по
элементов и включает все эти списки в новый список. Используйте
в качестве параметра функции.

Напишите функцию, вычисляющую факториал числа, передаваемого в качестве параметра.

Написать функцию умножения, принимающую от одного до трех параметров. Функцию вызвать с приведенными ниже аргументами. Для случая a4 выбрать 3 первых и 3 последних значения.

a1 = (15, 10 ,5)
a2 = (3, 1)
a3 = [2, 35, 55]
a4 = (5, 10 ,15, 20)
Напишите функцию, которая для заданного радиуса r вычисляет площадь круга и длину окружности. Функция возвращает кортеж из 2 значений.

Реализовать функции для выполнения четырех арифметических операций, преобразующих целые числа в целое число. Создать словарь с функциями исоответствующими им симв олами операций. Для двух заранее заданных целых чисел (например, 25 и 4) выполнить выбранную пользователем арифметическую операцию.

Написать калькулятор для строковых выражений вида '<число> <операция> <число>', где <число> - целое число, например 113, <арифметическая операция> - одна из операций +,-,*,//,%,^. Пример: calc('13 - 5') -> 8

Написать функцию, которая преобразует целое число (от 0 до 999) из текстового представления на русском языке в число типа int. Пример: to_int('тридцать три') -> 33

Написать функцию, которая преобразует целое число (от 0 до 999) из текстового представления на русском языке в число типа int. И сообщает об ошибках (выводит на экран описание типа ошибки и возвращает число -1). Пример: to_int('тридцать три') -> 33 Пример:

to_int("сто сорок тридцать два сто") -> -1 Вывод: тридцать - некорректное расположение в числе.

4. Выполните задания в Jupyter Notebook

Выполните задания

Создайте два множества
и
, добавив в
10 случайных целых чисел от 1 до 20, а в
– 5 таких же чисел. Для созданных множеств:
найдите их объединение, пересечение, разность, множество элементов, которые не входят в пересечение;
проверьте, является ли
подмножеством
и наоборот.
Создайте 2 списка из 5 случайных целых чисел от 1 до 10.
определите, сколько различных чисел содержат списки;
определите, сколько различных чисел содержится одновременно как в первом списке, так и во втором,
выведите все числа, которые входят как в первый, так и во второй список в порядке возрастания,
удалите из первого списка числа, входящие во второй список.
Для заданного натурального
найти все простые числа, которые меньше или равны
n, используя алгоритм, который называют «решето Эратосфена»:
пусть
– строимое множество простых чисел и
– множество, называемое решетом, в начале работы
– пустое множество,
– множество всех целых чисел от 2 до
;
на каждом шаге алгоритма наименьший элемент
помещается в
, а из
удаляются все числа, кратные этому элементу;
алгоритм заканчивает работу при пустом
.
Участники олимпиады решали 3 задачи. Известны фамилии тех, кто решил первую, вторую и третью задачи (для каждой задачи отдельный список). Найдите и выведите на экран фамилии тех, кто
решил хотя бы одну задачу (любую);
решил все задачи;
решил ровно 1 задачу (любую);
решил ровно 2 задачи (любые);
решил не больше 2 задач (любых).
Каждый участник международной конференции указал, какими языками он владеет (для хранения этой информации используйте словарь). Предполагается, что конференция проводится на русском языке, и выполняется синхронный перевод на английский. Определите:
есть ли язык, на котором разговаривают все участники;
фамилии участников, которые не владеют русским языком;
на какие языки еще нужен перевод (язык учитывается, если участник не знает русского и английского), и сколько человек знают этот язык.

5. Выполните задания в Jupyter Notebook

Выполните задания

1. Создайте словарь food продуктов питания, ключами которого являются наименования, а значениями цены. Исходные данные для словаря:

titles=('apples,grape,bananas,watermelons,tomatoes'.split(','))
prices=(20,25,15,11,15)

2. Удалите из словаря food арбузы (watermelons) и добавьте дыни (melons) по той же цене.

3. Все цены снижены на 15%. Обновите словарь продуктов.

4. Поменяйте местами цены и названия, пусть ключами словаря станут цены, а значениями - названия. Изменилось ли при этом количество элементов словаря? Если да, то объясните, почему.

5. Напишите функцию, возвращающую максимальную степень некоторого натурального числа
(например
), меньшую данного натурального числа
. Пусть
и
будут параметрами функции.

6. Создайте функцию, возвращающую ряд Фибоначчи до определенного натурального числа.

7. Создайте рекурсивную функцию, возвращающую ряд Фибоначчи до определенного натурального числа

8. Создайте функцию, которая проверяет, является ли число членом ряда Фибоначчи.

9. Студентка Александра родилась 11 июня 2003 года в 12 часов дня. Рассчитать время в минутах с момента рождения.

6. Выполните задания в Jupyter Notebook

Выполните задания

1. Дано: списки a,b=[1,1,1],[3,3,3]

Создать одномерные массивы A и B из списков a и b.

Вычислить выражение

2. Средствами NumPy рассчитать произведения четных чисел от 2 до 20 на ближайшие к ним бОльшие нечетные числа

3. Сгенерировать двумерный массив arr размерности (4, 7), состоящий из случайных действительных чисел, равномерно распределенных в дипазоне от 0 до 20. Нормализовать значения массива так, что после нормализации максимальный элемент масива будет равен 1.0, минимальный 0.0. Значения округлить (с помощью np.round()) до четырех десятичных знаков.

4. Имеется массив my_array

```
my_array = np.array([[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]])

Напишите код, с помощью которого можно извлечь из него центральный фрагмент размером 3 x 3, с числами 7, 8, 9, 12, 13, 14, 17, 18, 19.

5. Создайте массив my_sin, состоящий из синусов элементов массива my_array. Посчитайте, чему равна сумма элементов полученного массива. Ответ округлите до трёх цифр после запятой.

6. Создать матрицу из 20 случайных целых чисел от 0 до 100. Получить второе сверху значение в матрице. Определить индекс этого значения.

7. Задать два двухмерных массива ar1 и ar2 размерности (4, 7), состоящих из случайных целых чисел в пределах от 0 до 10. Построить двухмерный массив размерности (4, 7), каждый элемент которого представляет собой максимум (минимум, среднее) из двух значений, находящихся на аналогичной позиции в массивах ar1, ar2.

8. Имеется массив great_array

```
first = [x**(1/2) for x in range(100)]
second = [x**(1/3) for x in range(100, 200)]
third = [x/y for x in range(200,300,2) for y in [3,5]]
great_array = np.array([first, second, third]).T
```

Чему равна сумма элементов массива great_array, значение которых больше 50?

9. Создайте три массива: массив четных чисел от 2 до 20, массив ближайших к числам первого массива бОльших нечётных чисел, массив векторного произведения первого массива на второй. Соедините все массивы в один двумерный массив, расположив исходные массивы в столбцах.

$$-((B - A) * (B/2))$$

7. Выполните задания в Jupyter Notebook

Выполните задания

1. Имеется серия, содержащая числа от 10 до 1000. Какое число содержит элемент с индексом 175? Вычислите разность элементов с индексами 900 и 16.

2. Мебельная фабрика выпускает продукцию наименований: диваны, кровати, трюмо. Создайте таблицу DataFrame, содержащую: количество призведенной продукции каждого наименования, остаток на складе фабрики. Наименования разместите в строках, а количество и остаток в столбцах.

3. Разделите столбец car таблицы market на два столбца: mark (марка) и model (модель).

```
market = pd.DataFrame({'car':['Lada Granta','Lada Vesta','Kia Rio','Hyunday Creta'],
                       'year':[2019,2018,2010,2015],
                       'condition':[75,104,20,90], 'km':[130000,210000,75000,94000]})
```

4. Имеются исторические данные о цене акции в виде таблицы DataFrame.

```
history=pd.DataFrame([[150,20,140,140],[140,31,152,150],
                      [147,17,180,164],[129,33,150,167]],
                      columns=['Open','Low','High','Close'],
                      index=['2020-12-07','2020-12-08','2020-12-09','2020-12-10'])
```

Найдите минимальную и максимальную цены закрытия, стандартное отклонение цены

открытия, среднее значение самой низкой (Low) цены и медиану самой высокой (High) цены.

5. Загрузите этот набор данных в таблицу DataFrame. Какова сумма валовых продаж (GrossSales) товаров, произведеных Amarilla для малого бизнеса?

6. Определите среднюю стоимость производства (ManufacturingPrice) товаров, выпущенных для Мексики стоимостью (SalePrice) выше, чем стоимость 70% товаров.

7. Найдите минимальную цену производства товара правительственого сегмента, для которого цена продажи ниже цены производства.

8. Какова средняя цена товара, стоимость производства которого ниже средней?

9. Во сколько раз средняя себестоимость (COGS) товаров промышленного сегмента (Enterprize) отличается от средней себестоимости товаров малого бизнеса?

10. Какова суммарная прибыль всех сделок в правительственном сегменте, заключенных компаниями Carretera и Paseo?

11. Сколько разных систем скидок представлено в таблице?

12. Какая из систем скидок применяется реже всего?

13. Сколько раз представлены в выборке разные страны?

14. Разбейте значения столбца себестоимости (COGS) для Канады на шесть интервалов.

15. Разбейте значения столбца себестоимости (COGS) для Канады на шесть интервалов.

$$-((B-A)*(B/2))$$

Раздел 2. Разведочный анализ данных

Форма контроля/оценочное средство: Кейс-задание

Вопросы/Задания:

1. Имеются наборы данных. Необходимо в Jupyter Notebook реализовать программный код для выполнения следующих заданий.

1. Создайте Series из последовательности 15 значений, равномерно разбивающих отрезок [0, 20] (воспользуйтесь функцией linspace)
Определите отношение элементов полученной серии к их предыдущим элементам (*).
В результате необходимо получить среднее полученного вектора, оставив в нём только те значения, которые не более чем 1.5 (**).
Выберите из ответов тот, который максимально близок к полученному (с точки зрения абсолютной разницы).

2. Выберите все верные ответы касательно следующих четырех Series: - pd.Series('abcde'); (1) - pd.Series(['abcde']); (2) - pd.Series(list('abcde')); (3) - pd.Series("abcde"); (4)

Пояснения:
(*)

функция list: в строке каждый символ - это отдельный элемент для list
квадратные скобки: в квадратных скобках списку передается множество

3. По клиенту получены зашумленные данные (объект s типа Series) по его транзакциям.

Для заданного ниже объекта s проделайте следующее:

Создайте новый Series, значения которого совпадают со значениями s, а индексы - целочисленные значения от 2 до 12, не включая 12.
Выберите из s элементы с индексами 3 и 5, после чего просуммируйте их, сохранив результат (1).
Выберите из s только целочисленные элементы и вычислите их дисперсию (2). (*)
Все полученные результаты округлите до 2-х знаков после запятой.

4. Сгенерируйте Series из 100 значений нормально распределённой СВ (np.random.normal с дефолтными параметрами - нулевым средним и единичной дисперсией).
Возведите каждое значение серии в 3 степень, а значения индекса увеличьте в 3 раза.
Ответьте на следующие вопросы через запятую (без пробелов) (*)
Выведите сумму элементов, строго меньших 2.6, имеющих нечётные значения индекса.
Выведите количество значений серии меньше нуля.
Пояснения:
(*) Если получились ответы 3, 4.32, то в форму необходимо внести их в виде "3,4.32". То есть вещественные числа необходимо разделять точками. Не забудьте про фиксированный seed (его менять не нужно)!

Определенное значение seed нужно, чтобы ответы у всех выполняющих это задание были одинаковые и их можно было проверить (так как генерируются одинаковые series).
Следует внимательнее использовать [ ] для выбора данных по нескольким условиям: либо выбирать данные последовательно, либо сразу по нескольким условиям, но через оператор &.
Отличие оператора and от оператора &: and - выводит последнее проверенное значение, & - выводит пересечение значений. Пример: s[_ & _].sum()

Для всех последующих заданий будем использовать обезличенные транзакционные банковские данные. Для этого считайте в переменные tr_mcc_codes, tr_types, transactions и gender_train из одноимённых таблиц из папки data. Для таблицы transactions используйте только первые n=1000000 строк. Обратите внимание на разделители внутри каждого из файлов - они могут различаться!

Таблица transactions.csv
Описание
Таблица содержит историю транзакций клиентов банка за один год и три месяца.

Формат данных
customer_id,tr_datetime,mcc_code,tr_type,amount,term_id
111111,15 01:40:52,1111,1000,-5224,111111
111112,15 15:18:32,3333,2000,-100,11122233
...
Описание полей
customer_id — идентификатор клиента;
tr_datetime — день и время совершения транзакции (дни нумеруются с начала данных);
mcc_code — mcc-код транзакции;
tr_type — тип транзакции;
amount — сумма транзакции в условных единицах со знаком; + — начисление средств клиенту (приходная транзакция), - — списание средств (расходная транзакция);
term_id — идентификатор терминала;

Таблица gender_train.csv
Описание
Данная таблица содержит информацию по полу для части клиентов, для которых он известен. Для остальных клиентов пол неизвестен.

Формат данных
customer_id,gender
111111,0
111112,1
...
Описание полей
customer_id — идентификатор клиента;
gender — пол клиента;
Таблица tr_mcc_codes.csv
Описание
Данная таблица содержит описание mcc-кодов транзакций.

Формат данных
mcc_code;mcc_description
1000;словесное описание mcc-кода 1000
2000;словесное описание mcc-кода 2000
...
Описание полей
mcc_code – mcc-код транзакции;
mcc_description — описание mcc-кода транзакции.
Таблица tr_types.csv
Описание
Данная таблица содержит описание типов транзакций.

Формат данных
tr_type;tr_description
1000;словесное описание типа транзакции 1000
2000;словесное описание типа транзакции 2000
...
Описание полей
tr_type – тип транзакции;
tr_description — описание типа транзакции;

5. В tr_types выберите произвольные 100 строк с помощью метода sample (указав при этом random_seed равный 242)
В полученной на предыдущем этапе подвыборке найдите долю наблюдений (стобец tr_description), в которой содержится подстрока 'плата' (в любом регистре). (*)
Выведите ответ в виде вещественного числа, округлённого до двух знаков после запятой, отделив дробную часть точкой в формате "123.45"

Пояснения:
(*) Строки "ПлатА за аренду", "ПлатАза аренду", "ПЛАТА" удовлетворяют условию, так как будучи переведёнными в нижний регистр содержат подстроку "плата".

6. Для поля tr_type датафрейма transactions посчитайте частоту встречаемости всех типов транзакций tr_type в transactions.
Из перечисленных вариантов выберите те, которые попали в топ-5 транзакций по частоте встречаемости.
Выберите все верные пункты:
- 1) Выдача наличных в АТМ Сбербанк России
- 2) Комиссия за обслуживание ссудного счета
- 3) Списание по требованию
- 4) Оплата услуги. Банкоматы СБ РФ
- 5) Погашение кредита (в пределах одного филиала)

- 6) Покупка. POS ТУ СБ РФ

7. В датафрейме transactions задайте столбец customer_id в качестве индекса.
Выделите клиента с максимальной суммой транзакции (то есть с максимальным приходом на карту). (*)
Найдите у него наиболее часто встречающийся модуль суммы приходов/расходов. (**)

8. Найдите максимальную разницу между медианами суммы транзакций, посчитанными при заданных ниже условиях по полю amount из таблицы transactions (*):
Медиана суммы транзакций
Медиана суммы транзакций по тем строкам, которые ни в одном из своих столбцов не содержат пустые значения
Медиана суммы транзакций по строкам, отсортированным по полю amount в порядке возрастания, и из которых удалены дублирующиеся по столбцам [mcc_code, tr_type] строки, причём при удалении соответстующих дублей остаются только последние из дублирующихся строк (keep='last')
Выведите ответ в виде вещественного числа, округлённого до двух знаков после запятой, отделив дробную часть точкой в формате "123.45"

Пояснения:
(*) Для вычисления максимальной разницы между значениями списка можно использовать функцию np.ptp
(**) Если в результате получились значения [1,3,5], то макимальная разница между ними 4 == 5-1.

2. Имеются наборы данных. Необходимо в Jupyter Notebook реализовать программный код для выполнения следующих заданий.

1. Соедините transactions с всеми остальными таблицами (tr_mcc_codes, tr_types, gender_train). Причём с customers_gender_train необходимо смёрджиться с помощью left join, а с оставшимися датафреймами - через inner. После получения результата таблицы gender_train, tr_types, tr_mcc_codes можно удалить. В результате соединения датафреймов должно получиться 999584 строки.

2. Определите модуль разницы между средними тратами женщин и мужчин (трата - отрицательное значение amount). (*)
Выведите ответ в виде вещественного числа, округлённого до двух знаков после запятой, отделив дробную часть точкой в формате "123.45"

Пояснения:
(*) Если в результате для мужчин получились значения [-1,-3,-5], а для женщин [-1,-2,-3],
то модуль разницы между средними арифметическими -3 и -2 будет равен 1.

(**) Обратите внимание, что для вычисления модуля разности точных знаний о том,
какой класс относится к мужчинам, а какой - к женщинам, пока не требуется.

(***) Округление не нужно производить отдельно по средним тратам женщин и мужчин, а только в самом конце, когда получите значение модуля разницы трат.

3. Создайте новый столбец - mcc_code+tr_type, сконкатенировав значения из соответствующих столбцов. (*)
Оставьте только наблюдения с отрицательным значением amount. Посчитайте дисперсию по категориям получившегося столбца mcc_code+tr_type, в которых количество наблюдений >= 10.
Определите отношение максимальной дисперсии к минимальной.
Выведите ответ в виде вещественного числа, округлённого до ближайшего целого в формате "123456" без дробной части.

Пояснения:
(*) Для конкатенации значений в столбцах можно использовать метод .astype(str) для серии и складывать соответствующие серии. Либо же применять apply к строкам датафрейма, прописывая логику преобразования и конкатенации значений внутри.

(**) Для одновременного подсчета количества наблюдений и дисперсии по категориям можно воспользоваться функцией .agg()

4. По всем типам транзакций рассчитайте максимальную сумму прихода на карту (из строго положительных сумм по столбцу amount) отдельно для мужчин и женщин (назовите ее "max_income"). Оставьте по 5 транзакций для мужчин и для женщин, наименьших среди всех транзакций по полученным значениям "max_income". (*)
Выделите среди них те, которые встречаются одновременно и у мужчин, и у женщин:
Покупка. POS ТУ СБ РФ
Списание после проведения претензионной работы
Плата за получение наличных. Россия
Перевод на карту/ с карты через АТМ (со взиманием комиссии с отправителя) по счету в овердрафте
Плата за получение наличных в АТМ. Россия +
Наличные. Зарубеж. банк
Возврат покупки. POS ТУ Россия

5. Выделите из поля tr_datetime относительный день tr_day (первое число до точного времени). (*)
Отфильтруйте строки таким образом, чтобы оставить только те транзакции, у которых в соответствующий относительный день tr_day количество уникальных MCC кодов при транзакциях было больше 75 (можно воспользоваться функцией nunique())
Сгруппируйте полученный отфильтрованный датафрейм по MCC коду и полу, после чего, пронализировав результат, выберите верные варианты ответов ниже (**):
gender == 0 - женщины, gender == 1 - мужчины ++++
gender == 1 - женщины, gender == 0 - мужчины
Абсолютное значение медианы с типом "Флористика" (расходов/приходов) у мужчин выше той же медианы у женщин
Абсолютное значение медианы женских трат (расходов/приходов) на ценные бумаги выше мужских
Абсолютное значение медианы женских трат (расходов/приходов) в категории "Бары, коктейль-бары, дискотеки, ночные клубы и таверны — места продажи алкогольных напитков" ниже мужских
Пояснения:
(*) Для того, чтобы выделить всё, что стоит до первого пробела, можо использовать строковые методы для датафрейма - .str.split(), например. Либо же реализовывать логику выделения подстроки с помощью метода apply.
(**) Понять, какой класс к какому типу транзакций (мужские/женские) относится можно, если поизучать типичные для мужчин/женщин категории и сравнить средние/медианы расходов и/или приходов в них.

6. Разбейте расходы (отрицательные значения сумм) на 5 бакетов amount_bucket равного объёма (с помощью pd.qcut), разбив все траты на категории 'Very High', 'High', 'Middle', 'Low', 'Very Low'. (*)
Оставшиеся неотрицательные траты отнесите к категории 'Income'. (воспользуйтесь функцией cat.add_categories('Income').fillna('Income') для того, чтобы добавить новую категорию 'Income' к категориям 'Very High', 'High', 'Middle', 'Low', 'Very Low', а затем заполните пустые значения новой категорией).
Из поля tr_datetime выделите час tr_hour, в который произошла транзакция, как первые 2

цифры до ":". (**)
После этого постройте сводную таблицу, значениями в которой является пол gender, индексы - tr_hour, столбцы - amount_bucket.
Отрисуйте полученные результаты, передав их в функцию plot_pivot_table, расположенную ниже.
Выберите верные ответы на вопросы ниже.
1.0.0.7. Вопросы:
- 1) Ночные поступления денег (01-05 часов) в более чем 85% случаев являются мужскими.
- 2) Посмотрев на долю мужчин в поступлениях средств (Income), можно сделать вывод, что количество поступлений средств женщинам в целом больше, чем мужчинам.
- 3) Самые низкие траты в 3 часа ночи осуществляются в более 70% случаев женщинами.
- 4) Существуют особые часы в мелких тратах, когда женщины тратят намного больше мужчин (>80%)
- 5) Посмотрев на долю мужчин в максимальных тратах средств (Very High), можно сделать вывод, что количество высоких трат в каждый возможный час мужчин больше, чем у женщин.
Пояснения:
(*) Обратите внимание, что в категории Very High Должны оказаться максимальные по модулю отрицательные транзакции.
(**) Например, для строки "0 10:23:26" час будет равен 10, а для строки "6 07:08:31"- 07. Можно воспользоваться функциями str.split() или str.find() и функцией .apply(lambda x: x[])

7. Измените тип поля tr_day на int.
Выберите из transactions все MCC коды, которые встретились в выборке более чем 60000 раз.
Сгруппируйте отфильтрованный датафрейм по дню и MCC-коду, получая средние значения суммы amount.
Далее отрисуйте зависимость средних сумм (может пригодится метод unstack()) по каждому из MCC-кодов по дням.
Выберите верные ответы на вопросы ниже.
1.0.0.9. Вопросы:
- 1) 2 из полученных MCC-кодов связаны с финансовыми институтами
- 2) 2 MCC кода, связанные со снятием наличности имеют в целом разные знаки (в одном случае почти везде - траты, в другом - пополнения)
- 3) Бакалейные магазины обладают максимальными средними тратами среди выбранных MCC-кодов
- 4) Денежные переводы имеют как минимум 3 явных минимума средних
- 5) Категория "Звонки с использованием телефонов, считывающих магнитную ленту" имеет визуально очень большую дисперсию.

Раздел 3. Проверка статистических гипотез. Принятие решений

Форма контроля/оценочное средство: Задача

Вопросы/Задания:

1. Решите задачи используя библиотеку stats пакета scipy

Визуализируйте полученные результаты применяя библиотеку matplotlib
Оформите результаты в виде интерактивных блокнотов

Цель: изучить методы проверки статистических гипотез
Ход выполнения работы:

Решите задачи используя библиотеку stats пакета scipy
Визуализируйте полученные результаты применяя библиотеку matplotlib
Оформите результаты в виде интерактивных блокнотов

Задание для выполнения.
1. Завод производитель подшипников заявляет, что изготовленные на станках металлические элементы для подшипников, имеют средний диаметр 10 мм. Используя односторонний критерий с уровнем значимости α=0,05, проверить эту гипотезу. При проверке гипотезы необходимо учесть, что была произведена выборка из n=16 шариков, где среднее значение диаметра равно 10,3 мм, а дисперсия известна и равна 1 мм.

2. Производитель конфет заявляет, что средний вес коробки конфет составляет 100 г. Из партии извлечена выборка из n=10 коробок и взвешена. Вес каждой коробки соответствует таблице вариантов. Не противоречит ли это утверждению продавца? Используя уровень значимости α=0,001. Вес коробок конфет распределен нормально.

3. Произведены n=7 независимых измерений, в результате которых найдено, что \overline{x}=82,48 мм, а S=0,08. Предположив, что ошибки измерения имеют нормальное распределение проверить с использованием уровня значимости α=0,05 гипотезу H0:σ2=0,01 мм2. против конкурирующей гипотезы H0:σ2=0,005. В ответе записать разность между фактическим и табличным значениями выборочной характеристики.

4. Стратегия финансовой организации «Не обманешь!» не инвестирует в ценные бумаги если дисперсия годовой доходности более чем 0,04. Произведена выборка из n=52 наблюдений по активу А показала, что выборочная дисперсия ее доходности равна 0,045. Узнать, допустимы ли для данной финансовой организации инвестиционные вложения в актив А на уровне значимости: а) 0,05; б) 0,01.

5. Фирма «Спам» рассылает рекламные буклеты возможным заказчикам. Как показал опыт, вероятность того, что организация получившая буклет, закажет рекламируемое изделие, равна 0,08. Фирма разослала 1000 буклетов новой, улучшенной, формы и получила 100 заказов. На уровне значимости 0,05 выяснить, можно ли считать, что новая форма рекламы существенно лучше прежней.

6. Медицинский препарат «Огурчик» снимает похмельный синдром у 80% пациентов. Новый препарат «Огурчик NEW», разработанный для тех же целей, помог 90 пациентам из первых 100 применявших препарат. Можно ли на уровне значимости a = 0,05 считать, что новый препарат лучше? А на уровне a = 0,01?

7. Предполагается, что добавление специальных химических веществ в воду уменьшит ее жесткость. По оценке жесткости воды до и после добавления специальных веществ по 40-ка и 50-ти пробам соответственно получим средние значения жесткости (в стандартных единицах), равные 4,0 и 0,8. Дисперсия измерений в обоих случаях предполагается равно 0,25. Подтверждают ли эти результаты ожидаемый эффект? Принять α=0,05. Контролируемая величина имеет нормальное распределение.

8. Производительность каждого из перерабатывающих станков А и В составила (в кг вещества за час работы)

Можно ли считать производительность станков А и В одинаковой в предложении, что обе выборки получены из нормально распределенных генеральных совокупностей, при уровне значимости a = 0,1?

9. Перед наладкой станка была измерена точность изготовления 10 прокладок и найдено значение оценки дисперсии диаметра s21=9,6 мкм2. После наладки подверглись контролю еще 15 прокладок и получено новое значение оценки дисперсии s22=5,7 мкм2. Можно ли считать, что в результате наладки станка точность изготовления деталей увеличилась? Принять α=0,05.

10. При уровне значимости α=0,1 проверить гипотезу о равенстве дисперсий двух нормально распределенных случайных величин X и Y на основе выборочных данных при альтернативной гипотезе H1:σ2x≠σ2y.

11. Из 200 задач первого раздела курса «Анализа данных», предложенных для решения в лабораторных работах, студенты решили 130, а из 300 задач второго раздела студенты решили 120. Можно ли при α=0,01 утверждать, что первый раздел курса «Анализа данных» студенты усвоили лучше, чем второй.

12. Была проведена выборочная проверка надежности высокотехнологичной продукции 2-х производителей. В результате проверки были получены следующие результаты: в течения месяца после продажи в 15 из 200 технологических продуктов производителя А обнаружены дефекты, тогда как среди 400 продуктов производителя В - 8% оказались дефектами. Существенны ли различия в надежности продукции производителей А и В? Уровень

значимости принять равным 0,01.

Форма контроля/оценочное средство: Расчетно-графическая работа

Вопросы/Задания:

1. В этом интерактивном блокноте:

оценим ошибки первого и второго рода для теста о доле с помощью симуляций (его мы применяли для Джеймса Бонда)
посмотрим на то, как можно рассчитать число наблюдений необходимое для конкретных величин ошибок
Менеджер Алексей хочет проверить правда ли Джеймс Бонд отличает взболтанный мартини от смешанного. Алексей полагает, что если Бонд правда умеет различать напитки, то размер эффекта должен быть как минимум 0.2. Алексей хотел бы получить ошибки первого и второго рода равные 1%. Сколько наблюдений ему нужно? Посчитайте ошибки первого и второго рода. Посмотрим на то как ошибки зависят друг от друга в зависимости от выбора критического значения.

7. Оценочные материалы промежуточной аттестации

Пятый семестр, Экзамен

Контролируемые ИДК: ОПК-4.1 ОПК-4.2 ОПК-4.3 ОПК-4.4

Вопросы/Задания:

1. Этапы анализа данных

Этапы анализа данных

2. Работа с числовыми признаками

Работа с числовыми признаками

3. Работа с категориальными признаками

Работа с категориальными признаками

4. Предобработка данных

Предобработка данных

5. Описательные статистики

Описательные статистики

6. Борьба с выбросами (правило 3-х сигм, интерквартильных размах)

Борьба с выбросами (правило 3-х сигм, интерквартильных размах)

7. Зависимые и независимые случайные величины

Зависимые и независимые случайные величины

8. Виды корреляций

Виды корреляций

9. Масштабирование и нормирование данных

Масштабирование и нормирование данных

10. Метод моментов

Метод моментов

11. Дельта метод

Дельта метод

12. Несмещенная, Состоятельная, Эффективная оценки

Несмещенная, Состоятельная, Эффективная оценки

13. Асимптотический доверительный интервал для среднего

Асимптотический доверительный интервал для среднего

14. Асимптотический доверительный интервал для долей

Асимптотический доверительный интервал для долей

15. Асимптотический доверительный интервал для разности

Асимптотический доверительный интервал для разности

16. Точные доверительные интервалы для нормальных выборок

Точные доверительные интервалы для нормальных выборок

17. Схема проверки гипотез

Схема проверки гипотез

18. Схема АВ тестирования

Схема АВ тестирования

19. Фильтрация по условию

Фильтрация по условию

20. Агрегация и группировка

Агрегация и группировка

21. Объединение таблиц

Объединение таблиц

22. Numpy работа с числовыми массивами

Numpy работа с числовыми массивами

23. Визуализация данных

Визуализация данных

24. Pandas: Предобработка данных

Pandas: Предобработка данных

25. Pandas: Работа с индексами

Pandas: Работа с индексами

26. Работа с датой и временем

Работа с датой и временем

8. Материально-техническое и учебно-методическое обеспечение дисциплины

8.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Анализ данных: учеб. пособие / Краснодар: КубГАУ, 2018. - 126 с. - 978-5-00097-530-5. - Текст: электронный. // : [сайт]. - URL: https://edu.kubsau.ru/mod/resource/view.php?id=5007 (дата обращения: 02.05.2024). - Режим доступа: по подписке

2. ПАВЛОВ Д. А. Анализ данных: метод. рекомендации / ПАВЛОВ Д. А.. - Краснодар: КубГАУ, 2020. - 31 с. - Текст: электронный. // : [сайт]. - URL: https://edu.kubsau.ru/mod/resource/view.php?id=8053 (дата обращения: 02.05.2024). - Режим доступа: по подписке

Дополнительная литература

1. Новикова О. А. Анализ данных. Часть 1: Учебное пособие / Новикова О. А., Андрианова Е. Г.. - Москва: РТУ МИРЭА, 2020. - 162 с. - Текст: электронный. // RuSpLAN: [сайт]. - URL: https://e.lanbook.com/img/cover/book/167597.jpg (дата обращения: 21.02.2024). - Режим доступа: по подписке

2. Котиков П. Е. Анализ данных: учебно-методическое пособие / Котиков П. Е.. - Санкт-Петербург: СПбГПМУ, 2019. - 48 с. - 978-5-907184-46-6. - Текст: электронный. // RuSpLAN: [сайт]. - URL: https://e.lanbook.com/img/cover/book/174498.jpg (дата обращения: 21.02.2024). - Режим доступа: по подписке

3. Маккинли,, Уэс Python и анализ данных / Уэс Маккинли,; перевод А. Слинкина. - Python и анализ данных - Саратов: Профобразование, 2019. - 482 с. - 978-5-4488-0046-7. - Текст: электронный. // IPR SMART: [сайт]. - URL: https://www.iprbookshop.ru/88752.html (дата обращения: 20.02.2024). - Режим доступа: по подписке

8.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся

Профессиональные базы данных
Не используются.

Ресурсы «Интернет»

1. https://znanium.com/ - Znanium.com
2. https://edu.kubsau.ru/ - Образовательный портал КубГАУ
3. http://www.iprbookshop.ru/ - IPRbook

8.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине

Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине позволяют:
– обеспечить взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети «Интернет»;
– фиксировать ход образовательного процесса, результатов промежуточной аттестации по дисциплине и результатов освоения образовательной программы;
– организовать процесс образования путем визуализации изучаемой информации посредством использования презентаций, учебных фильмов;
– контролировать результаты обучения на основе компьютерного тестирования.
Перечень лицензионного программного обеспечения:
1 Microsoft Windows - операционная система.
2 Microsoft Office (включает Word, Excel, Power Point) - пакет офисных приложений.
Перечень профессиональных баз данных и информационных справочных систем:
1 Гарант - правовая, https://www.garant.ru/
2 Консультант - правовая, https://www.consultant.ru/
3 Научная электронная библиотека eLibrary - универсальная, https://elibrary.ru/
Доступ к сети Интернет, доступ в электронную информационно-образовательную среду университета.

Перечень программного обеспечения
(обновление производится по мере появления новых версий программы)
Не используется.

Перечень информационно-справочных систем
(обновление выполняется еженедельно)
Не используется.

8.4. Специальные помещения, лаборатории и лабораторное оборудование

Компьютерный класс
226гл
Интерактивная панель Samsung - 1 шт.
Персональный компьютер HP 6300 Pro SFF/Core i3-3220/4GB/500GB/NoODD/Win7Pro - 1 шт.
Сплит-система LS-H12KPA2/LU-H12KPA2 - 1 шт.

9. Методические указания по освоению дисциплины (модуля)

Учебная работа по направлению подготовки осуществляется в форме контактной работы с

преподавателем, самостоятельной работы обучающегося, текущей и промежуточной аттестаций, иных формах, предлагаемых университетом. Учебный материал дисциплины структурирован и его изучение производится в тематической последовательности. Содержание методических указаний должно соответствовать требованиям Федерального государственного образовательного стандарта и учебных программ по дисциплине. Самостоятельная работа студентов может быть выполнена с помощью материалов, размещенных на портале поддержки Moodl.

Методические указания по формам работы

Лекционные занятия

Передача значительного объема систематизированной информации в устной форме достаточно большой аудитории. Дает возможность экономно и систематично излагать учебный материал. Обучающиеся изучают лекционный материал, размещенный на портале поддержки обучения Moodl.

Лабораторные занятия

Практическое освоение студентами научно-теоретических положений изучаемого предмета, овладение ими техникой экспериментирования в соответствующей отрасли науки. Лабораторные занятия проводятся с использованием методических указаний, размещенных на образовательном портале университета.

Описание возможностей изучения дисциплины лицами с ОВЗ и инвалидами

Для инвалидов и лиц с ОВЗ может изменяться объём дисциплины (модуля) в часах, выделенных на контактную работу обучающегося с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающегося (при этом не увеличивается количество зачётных единиц, выделенных на освоение дисциплины).

Фонды оценочных средств адаптируются к ограничениям здоровья и восприятия информации обучающимися.

Основные формы представления оценочных средств – в печатной форме или в форме электронного документа.

Формы контроля и оценки результатов обучения инвалидов и лиц с ОВЗ с нарушением зрения:

– устная проверка: дискуссии, тренинги, круглые столы, собеседования, устные коллоквиумы и др.;

– с использованием компьютера и специального ПО: работа с электронными образовательными ресурсами, тестирование, рефераты, курсовые проекты, дистанционные формы, если позволяет острота зрения - графические работы и др.;

– при возможности письменная проверка с использованием рельефно-точечной системы Брайля, увеличенного шрифта, использование специальных технических средств (тифлотехнических средств): контрольные, графические работы, тестирование, домашние задания, эссе, отчеты и др.

Формы контроля и оценки результатов обучения инвалидов и лиц с ОВЗ с нарушением слуха:

– письменная проверка: контрольные, графические работы, тестирование, домашние задания, эссе, письменные коллоквиумы, отчеты и др.;

– с использованием компьютера: работа с электронными образовательными ресурсами, тестирование, рефераты, курсовые проекты, графические работы, дистанционные формы и др.;

– при возможности устная проверка с использованием специальных технических средств (аудиосредств, средств коммуникации, звукоусиливающей аппаратуры и др.): дискуссии, тренинги, круглые столы, собеседования, устные коллоквиумы и др.

Формы контроля и оценки результатов обучения инвалидов и лиц с ОВЗ с нарушением опорно-двигательного аппарата:

– письменная проверка с использованием специальных технических средств (альтернативных средств ввода, управления компьютером и др.): контрольные, графические работы, тестирование, домашние задания, эссе, письменные коллоквиумы, отчеты и др.;
– устная проверка, с использованием специальных технических средств (средств коммуникаций): дискуссии, тренинги, круглые столы, собеседования, устные коллоквиумы и др.;
– с использованием компьютера и специального ПО (альтернативных средств ввода и управления компьютером и др.): работа с электронными образовательными ресурсами, тестирование, рефераты, курсовые проекты, графические работы, дистанционные формы предпочтительнее обучающимся, ограниченным в передвижении и др.
Адаптация процедуры проведения промежуточной аттестации для инвалидов и лиц с ОВЗ.
В ходе проведения промежуточной аттестации предусмотрено:
– предъявление обучающимся печатных и (или) электронных материалов в формах, адаптированных к ограничениям их здоровья;
– возможность пользоваться индивидуальными устройствами и средствами, позволяющими адаптировать материалы, осуществлять приём и передачу информации с учетом их индивидуальных особенностей;
– увеличение продолжительности проведения аттестации;
– возможность присутствия ассистента и оказания им необходимой помощи (занять рабочее место, передвигаться, прочитать и оформить задание, общаться с преподавателем).
Формы промежуточной аттестации для инвалидов и лиц с ОВЗ должны учитывать индивидуальные и психофизические особенности обучающегося/обучающихся по АОПОП ВО (устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п.).
Специальные условия, обеспечиваемые в процессе преподавания дисциплины студентам с нарушениями зрения:
– предоставление образовательного контента в текстовом электронном формате, позволяющем переводить плоскопечатную информацию в аудиальную или тактильную форму;
– возможность использовать индивидуальные устройства и средства, позволяющие адаптировать материалы, осуществлять приём и передачу информации с учетом индивидуальных особенностей и состояния здоровья студента;
– предоставление возможности предкурсового ознакомления с содержанием учебной дисциплины и материалом по курсу за счёт размещения информации на корпоративном образовательном портале;
– использование чёткого и увеличенного по размеру шрифта и графических объектов в мультимедийных презентациях;
– использование инструментов «лупа», «прожектор» при работе с интерактивной доской;
– озвучивание визуальной информации, представленной обучающимся в ходе занятий;
– обеспечение раздаточным материалом, дублирующим информацию, выводимую на экран;
– наличие подписей и описания у всех используемых в процессе обучения рисунков и иных графических объектов, что даёт возможность перевести письменный текст в аудиальный;
– обеспечение особого речевого режима преподавания: лекции читаются громко, разборчиво, отчётливо, с паузами между смысловыми блоками информации, обеспечивается интонирование, повторение, акцентирование, профилактика рассеивания внимания;
– минимизация внешнего шума и обеспечение спокойной аудиальной обстановки;
– возможность вести запись учебной информации студентами в удобной для них форме (аудиально, аудиовизуально, на ноутбуке, в виде пометок в заранее подготовленном тексте);
– увеличение доли методов социальной стимуляции (обращение внимания, апелляция к ограничениям по времени, контактные виды работ, групповые задания и др.) на практических и лабораторных занятиях;
– минимизирование заданий, требующих активного использования зрительной памяти и зрительного внимания;
– применение поэтапной системы контроля, более частый контроль выполнения заданий для самостоятельной работы.
Специальные условия, обеспечиваемые в процессе преподавания дисциплины студентам с нарушениями опорно-двигательного аппарата (маломобильные студенты, студенты, имеющие

трудности передвижения и патологию верхних конечностей):
– возможность использовать специальное программное обеспечение и специальное оборудование и позволяющее компенсировать двигательное нарушение (коляски, ходунки, трости и др.);
– предоставление возможности предкурсового ознакомления с содержанием учебной дисциплины и материалом по курсу за счёт размещения информации на корпоративном образовательном портале;
– применение дополнительных средств активизации процессов запоминания и повторения;
– опора на определенные и точные понятия;
– использование для иллюстрации конкретных примеров;
– применение вопросов для мониторинга понимания;
– разделение изучаемого материала на небольшие логические блоки;
– увеличение доли конкретного материала и соблюдение принципа от простого к сложному при объяснении материала;
– наличие чёткой системы и алгоритма организации самостоятельных работ и проверки заданий с обязательной корректировкой и комментариями;
– увеличение доли методов социальной стимуляции (обращение внимания, аппеляция к ограничениям по времени, контактные виды работ, групповые задания др.);
– обеспечение беспрепятственного доступа в помещения, а также пребывания них;
– наличие возможности использовать индивидуальные устройства и средства, позволяющие обеспечить реализацию эргономических принципов и комфортное пребывание на месте в течение всего периода учёбы (подставки, специальные подушки и др.).
Специальные условия, обеспечиваемые в процессе преподавания дисциплины студентам с нарушениями слуха (глухие, слабослышащие, позднооглохшие):
– предоставление образовательного контента в текстовом электронном формате, позволяющем переводить аудиальную форму лекции в плоскопечатную информацию;
– наличие возможности использовать индивидуальные звукоусиливающие устройства и сурдотехнические средства, позволяющие осуществлять приём и передачу информации; осуществлять взаимообратный перевод текстовых и аудиофайлов (блокнот для речевого ввода), а также запись и воспроизведение зрительной информации;
– наличие системы заданий, обеспечивающих систематизацию вербального материала, его схематизацию, перевод в таблицы, схемы, опорные тексты, глоссарий;
– наличие наглядного сопровождения изучаемого материала (структурно-логические схемы, таблицы, графики, концентрирующие и обобщающие информацию, опорные конспекты, раздаточный материал);
– наличие чёткой системы и алгоритма организации самостоятельных работ и проверки заданий с обязательной корректировкой и комментариями;
– обеспечение практики опережающего чтения, когда студенты заранее знакомятся с материалом и выделяют незнакомые и непонятные слова и фрагменты;
– особый речевой режим работы (отказ от длинных фраз и сложных предложений, хорошая артикуляция; четкость изложения, отсутствие лишних слов; повторение фраз без изменения слов и порядка их следования; обеспечение зрительного контакта во время говорения и чуть более медленного темпа речи, использование естественных жестов и мимики);
– чёткое соблюдение алгоритма занятия и заданий для самостоятельной работы (называние темы, постановка цели, сообщение и запись плана, выделение основных понятий и методов их изучения, указание видов деятельности студентов и способов проверки усвоения материала, словарная работа);
– соблюдение требований к предъявляемым учебным текстам (разбивка текста на час¬ти; выделение опорных смысловых пунктов; использование наглядных средств);
– минимизация внешних шумов;
– предоставление возможности соотносить вербальный и графический материал; комплексное использование письменных и устных средств коммуникации при работе в группе;
– сочетание на занятиях всех видов речевой деятельности (говорения, слушания, чтения, письма, зрительного восприятия с лица говорящего).
Специальные условия, обеспечиваемые в процессе преподавания дисциплины студентам с

прочими видами нарушений (ДЦП с нарушениями речи, заболевания эндокринной, центральной нервной и сердечно-сосудистой систем, онкологические заболевания):
– наличие возможности использовать индивидуальные устройства и средства, позволяющие осуществлять приём и передачу информации;
– наличие системы заданий, обеспечивающих систематизацию вербального материала, его схематизацию, перевод в таблицы, схемы, опорные тексты, глоссарий;
– наличие наглядного сопровождения изучаемого материала;
– наличие чёткой системы и алгоритма организации самостоятельных работ и проверки заданий с обязательной корректировкой и комментариями;
– обеспечение практики опережающего чтения, когда студенты заранее знакомятся с материалом и выделяют незнакомые и непонятные слова и фрагменты;
– предоставление возможности соотносить вербальный и графический материал; комплексное использование письменных и устных средств коммуникации при работе в группе;
– сочетание на занятиях всех видов речевой деятельности (говорения, слушания, чтения, письма, зрительного восприятия с лица говорящего);
– предоставление образовательного контента в текстовом электронном формате;
– предоставление возможности предкурсового ознакомления с содержанием учебной дисциплины и материалом по курсу за счёт размещения информации на корпоративном образовательном портале;
– возможность вести запись учебной информации студентами в удобной для них форме (аудиально, аудиовизуально, в виде пометок в заранее подготовленном тексте);
– применение поэтапной системы контроля, более частый контроль выполнения заданий для самостоятельной работы;
– стимулирование выработки у студентов навыков самоорганизации и самоконтроля;
– наличие пауз для отдыха и смены видов деятельности по ходу занятия.

10. Методические рекомендации по освоению дисциплины (модуля)