

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет математики и компьютерных наук
Кафедра прикладной математики

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине
Статистический анализ в Python

Кафедра прикладной математики
факультета математики и компьютерных наук

Образовательная программа бакалавриата
01.03.05 – Статистика

Направленность (профиль) программы
Анализ больших данных

Форма обучения
Очная

Статус дисциплины: *входит в часть, формируемая участниками образовательных отношений ОПОП*

Махачкала, 2023

Фонд оценочных средств по дисциплине «Статистический анализ в Python» составлена в 2023 году в соответствии с требованиями ФГОС ВО бакалавриата по направлению подготовки 01.03.05 - статистика от 14.08.2020 г. № 1032

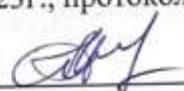
Разработчики:

1. кафедра прикладной математики, Гаджиева Т.Ю. к.ф.-м. н., доцент;

Фонд оценочных средств по дисциплине «Статистический анализ в Python» одобрен:
на заседании кафедры прикладной математики от «20» января 2023г., протокол №5

Зав. кафедрой  Кадиев Р.И.

на заседании Методической комиссии факультета математики и компьютерных наук от «25» января 2023г., протокол №4.

Председатель  Ризаев М.К.

Фонд оценочных средств «Статистический анализ нечисловой информации» согласован с учебно-методическим управлением

«20» февраля 2023г.

Начальник УМУ  Гасангаджиева А.Г.

**1. ПАСПОРТ
ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине
«Статистический анализ в Python»**

1.1. Основные сведения о дисциплине

Общая трудоемкость дисциплины составляет 3 зачетные единицы (108 академических часов).

Вид работы	Трудоемкость, академических часов		
	б семестр	__ семестр	всего
Общая трудоёмкость	108		108
Контактная работа:	32		32
Лекции (Л)	16		16
Практические занятия (ПЗ)			
Лабораторные занятия (ЛЗ)	16		16
Консультации			
Промежуточная аттестация (зачет, экзамен)	экзамен		
Самостоятельная работа	76		76
1. работа с лекционным материалом, с учебной литературой	5		5
2. опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	5		5
3. выполнение домашних заданий	10		10
4. подготовка к лабораторным работам, к практическим занятиям	10		10
5. подготовка к коллоквиуму	5		5
6. подготовка к контрольным работам	5		5
7. подготовка к экзамену	36		36

1.2. Требования к результатам обучения по дисциплине, формы их контроля и виды оценочных средств

*ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине «Статистический анализ в Python»*

№ п/п	Контролируемые модули, разделы (темы) дисциплины	Код контролируемой компетенции (или её части)	Оценочные средства		Способ контроля
			наименование	№№ заданий	
1	Модуль 1. Методы и алгоритмы аналитики статистических данных	ПК-5	Вопросы для собеседования	1-10	устно
		ПК-5		1	письменно

			Контрольные работы		
2	Модуль 2. Статистический анализ данных в Python	ПК-5	Вопросы для собеседования	11-21	устно
		ПК-5	Контрольные работы	2	письменно
3	Модуль 3. Экзамен	ПК-5	Вопросы для собеседования	22-35	устно
		ПК-5	Контрольные работы	3	письменно

1.3. Показатели и критерии определения уровня сформированности компетенций

№ п/п	Код компетенции	Уровни сформированности компетенции			
		Недостаточный	Удовлетворительный (достаточный)	Базовый	Повышенный
		Отсутствие признаков удовлетворительного уровня	Знать: Уметь: Владеть:	Знать: Уметь: Владеть:	Знать: Уметь: Владеть:
1	ПК-5	Не знает как разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Не умеет разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Не владеет: навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	Знает на достаточно хорошем уровне как разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Умеет на достаточно хорошем уровне как разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Владеет на достаточно хорошем уровне как навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	Знает на хорошем уровне как разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Умеет на хорошем уровне разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Владеет на хорошем уровне навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	Знает в совершенстве как разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Умеет в совершенстве разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Владеет в совершенстве навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.

2. КОНТРОЛЬНЫЕ ЗАДАНИЯ И ИНЫЕ МАТЕРИАЛЫ ОЦЕНКИ знаний, умений, навыков и (или) опыта деятельности, характеризующие этапы формирования компетенций в процессе освоения дисциплины «Статистический анализ в Python»

Контрольная работа Контрольная работа 1

Задача 1. Создать следующего вида исходный файл из восьми строк, содержащий в каждой строке значение исходной системы счисления, результирующей системы счисления и пяти различных целых положительных чисел, корректных для исходной системы счисления:

10 2 121 13 10 2018 5

Задача 2. Создать следующего вида исходный файл из десяти строк, содержащий в каждой строке ФИО, дату рождения, дату получения паспорта и баллы ЕГЭ по трём предметам:

Апельсинов А.А. 08.02.2000 17.03.2014 90 100 91

Виноградов В.В. 09.03.1999 15.04.2013 67 99 98

Яблоков Я.Я. 10.04.2000 19.05.2014 94 94 94

Морковкин М.М. 11.05.1999 17.06.2013 91 82 73

Задача 5. Импортируйте набор данных о пассажирах «Титаника». Проведите разведывательный анализ данных. Поставьте гипотезу. Проведите необходимые преобразования данных.

Задача 6. На основании поставленной гипотезы в задании 1 выберите модели для предсказания. Обоснуйте выбор. Постройте выбранные модели, выберите лучшую. Постройте ансамбль моделей. Улучшились ли предсказания? Оформите результаты в виде отчета.

Задача 7. Вариант проекта: соберите данные для анализа с сайта Федеральной службы государственной статистики. Проведите предварительную обработку данных. Поставьте задачу классификации. Проведите необходимые преобразования данных. Постройте модель. Оцените, насколько хорошо модель справляется с поставленной задачей. Опишите результат, назовите возможные причины плохой работы модели (в случае, если результаты неудовлетворительные).

Контрольная работа 2

Задача 1. Напишите программу Pandas для отображения всех записей файла REGIONS.

Задача 2. Напишите программу Pandas для отображения имени, фамилии, оклада и идентификатора менеджера, где идентификаторы менеджера равны нулю.

Задача 3. Напишите программу Pandas для отображения всех идентификаторов местоположений из файла местоположений.

Задача 4. Напишите программу Pandas для извлечения первых 7 записей из файла сотрудников.

Задача 5. Напишите программу Pandas, чтобы выбрать отдельный идентификатор отдела из файла сотрудников.

Задача 6. Напишите программу Pandas, в которой будут отображаться имя и фамилия, а также номер отдела для всех сотрудников, чья фамилия «Иванов».

Задача 7. Напишите программу Pandas для отображения имени, фамилии, оклада и идентификатора менеджера, где идентификаторы менеджера не равны нулю.

Задача 8. Напишите программу Pandas, чтобы отобразить имя, фамилию, оклад и номер отдела для тех сотрудников, для которых буква n является третьим символом в их имени.

Задача 9. Напишите программу Pandas для отображения имени, идентификатора задания, зарплаты и отдела для тех сотрудников, которые не работают в отделах 50, 30 и 80.

Критерии оценки:

- оценка «отлично» выставляется студенту, если верно и правильно выполнено 90%-100% заданий;
- оценка «хорошо» выставляется студенту, если верно и правильно выполнено 70%-80% заданий;
- оценка «удовлетворительно» выставляется студенту, если верно и правильно решено 50%-60% заданий, возможны некоторые исправления при решении;
- оценка «неудовлетворительно» выставляется студенту, если верно выполнено менее 50% заданий;

Примеры домашних заданий

1. Импортируйте набор данных о пассажирах “Титаника”. Проведите разведывательный анализ данных. Поставьте гипотезу. Проведите необходимые преобразования данных.

2. На основании поставленной гипотезы в ДЗ1 выберите модели для предсказания. Обоснуйте выбор. Постройте выбранные модели, выберите лучшую. Постройте ансамбль моделей. Улучшились ли предсказания? Оформите результаты в виде отчета.

3. Вариант проекта: соберите данные из социальной сети или с новостного сайта для последующего текстового анализа. Проведите предварительную обработку данных. Поставьте задачу классификации. Проведите необходимые преобразования данных. Постройте модель. Оцените, насколько хорошо модель справляется с поставленной задачей. Опишите результат, назовите возможные причины плохой работы модели (в случае, если результаты неудовлетворительные).

Критерии оценки:

- оценка «отлично» выставляется студенту, если верно и правильно выполнено 90%-100% заданий;
- оценка «хорошо» выставляется студенту, если верно и правильно выполнено 70%-80% заданий;
- оценка «удовлетворительно» выставляется студенту, если верно и правильно решено 50%-60% заданий, возможны некоторые исправления при решении;
- оценка «неудовлетворительно» выставляется студенту, если верно выполнено менее 50% заданий;

Вопросы для коллоквиумов, собеседования

МОДУЛЬ 1. Методы и алгоритмы аналитики статистических данных

1. Основы программирования на языке Python (refresher).
2. Математический аппарат (refresher). Введение в модуль NumPy.
3. Основы работы с Pandas. Разведывательный анализ данных
4. Особенности типов данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).
5. Визуализация данных. Представление результатов исследования
6. Работа с табличными данными в Pandas на примере данных Всемирного банка по продолжительности жизни и данных о публикационной активности словацких экономистов
7. Работа с текстовыми данными. Сбор данных из открытых источников.
8. Создание различных видов визуализаций на синтетических данных и тренировочных наборах данных.
9. Создание визуализаций для разведывательного анализа данных для набора данных о пассажирах “Титаника”.

МОДУЛЬ 2: Статистический анализ данных в Python

1. Анализ сетей
2. Введение в модуль NetworkX. Подготовка данных для анализа сетей. Примеры визуализации сетей на примере данных из сети Twitter.
3. Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.
4. Библиотека sklearn. Постановка гипотезы, преобразование данных и построение регрессионных моделей для задачи с пассажирами “Титаника”
5. Деревья решений. Случайный лес. Ансамбли моделей
6. Введение в нейронные сети.

Задание к лабораторной работе № 1 (ЛР1) (8 баллов)

«Программирование линейных алгоритмов»

Напишите программу для расчета по двум формулам. Подготовьте не менее пяти тестовых примеров. Предварительно выполните вычисления с использованием калькулятора или офисного приложения, например, Excel. Результаты вычисления по обеим формулам должны совпадать. Отсутствующие в языке функции выразите через имеющиеся.

$$1) y = \operatorname{tg}^2\left(\frac{x^2}{2} - 1\right) + \frac{2\cos(x - \pi/6)}{1/2 + \sin^2 \alpha}; \quad 2) y = 2 \frac{\log(3 + \sin(x))(3 - \cos(\pi/4 + 2x))}{1 + \operatorname{tg}^2(2x/\pi)}$$

Описание алгоритма

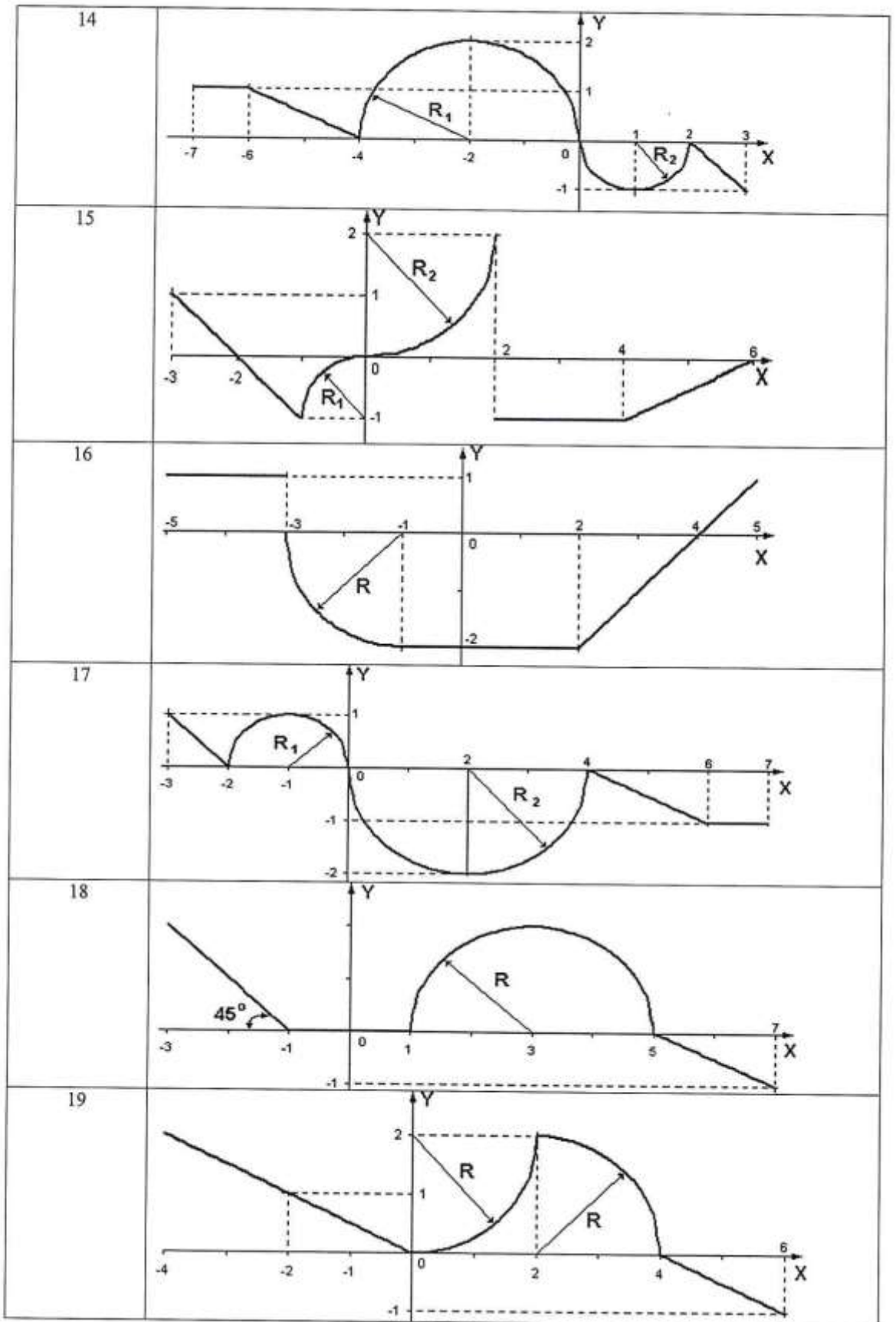
Для вычислений необходимо обеспечить ввод двух переменных x и a . Поскольку по условиям задачи их тип и точность представления не заданы, выберем для них вещественный тип (float). Для оптимизации записи выражения используем промежуточную переменную tmp .

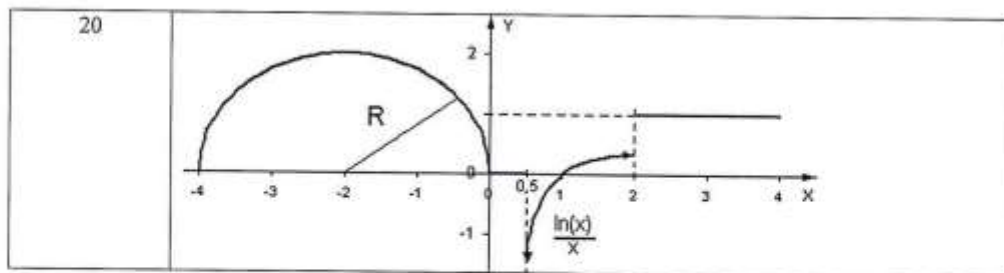
1. Ввести значения a и x , преобразовать к типу float.
2. Вычислить выражение 1.
3. Вывести результат вычисления.
4. Вычислить значение переменной tmp ;
5. Вычислить выражение 2.
6. Вывести результат вычисления.

Напишите программу для расчета по двум формулам. Подготовьте не менее пяти тестовых примеров. Предварительно выполните вычисления с использованием калькулятора или офисного приложения, например, Excel. Результаты вычисления по обеим формулам должны совпадать. Отсутствующие в языке функции выразите через имеющиеся.

Таблица 1. Варианты заданий Лабораторной работы 2

вариант	задание
1	
2	
3	
4	
5	
6	
7	





Задание к лабораторной работе № 3 (ЛР3) (8 баллов)

Введение в модуль NumPy. Основы работы с Pandas. Разведывательный анализ данных

Вариант 1

Импортировать модуль NumPy. Создать целочисленный двумерный массив 3x4. провести операции сложения с числом и поэлементное умножение на другой массив той же размерности. Вычислить сумму чисел.

Вариант 2

Импортировать модуль NumPy. Создать одномерный массив случайных чисел. Вычислить сумму чисел. Построить гистограмму.

Вариант 3

Импортировать модуль NumPy. Создать целочисленный двумерный массив 3x4 и двумерный целочисленный массив 4 x 2. Получить матричное произведение этих массивов. С помощью среза отбросить крайние столбцы из результирующего массива.

Вариант 4

Импортировать модуль NumPy. Создать одномерный массив чисел от 1 до 100 с шагом 5. Создать второй массив применением к первому функции $\sin()$. Вывести график.

Вариант 5

Скачать датасет по успеваемости в формате csv. <https://www.kaggle.com/spscientist/students-performance-in-exams> Импортировать модуль pandas. Загрузить данные. Отфильтровать только мужчин. Рассчитать средние баллы по предметам. Построить гистограмму.

Вариант 5

Скачать датасет по успеваемости в формате csv. <https://www.kaggle.com/spscientist/students-performance-in-exams> Импортировать модуль pandas. Загрузить данные. Отфильтровать тех, кто знает математику на 70 или больше баллов. Рассчитать средние баллы по предметам для мужчин и женщин. Построить боксплот.

Вариант 7

Скачать датасет по успеваемости в формате csv. <https://www.kaggle.com/spscientist/students-performance-in-exams> Импортировать модуль pandas. Загрузить данные. Выбрать 100 случайных студентов. Рассчитать средние баллы по предметам для мужчин и женщин. Построить боксплот.

Вариант 8

Скачать датасет по успеваемости в формате csv. <https://www.kaggle.com/spscientist/students-performance-in-exams> Импортировать модуль pandas. Загрузить данные.. Построить боксплот с разбиением на мужчин и женщин.

Вариант 9

Скачать датасет по успеваемости в формате csv. <https://www.kaggle.com/spscientist/students-performance-in-exams> Импортировать модуль pandas. Загрузить данные..Найти частоту значений пола и других признаков. Построить диаграммы рассеяния для всех предметов.

Вариант 10

Скачать датасет по успеваемости в формате csv. <https://www.kaggle.com/spscientist/students-performance-in-exams> Импортировать модуль pandas. Загрузить данные. Отфильтровать только тех, кто получает завтрак в школе. Найти все статические характеристики по всем предметам.

Задание к лабораторной работе № 4 (ЛР4) (8 баллов)

«Организация циклов»

Вариант 1

Вычислить и вывести на экран в виде таблицы значения функции, заданной графически (см. задание лабораторной работы № 2), на интервале от $X_{нач}$ до $X_{кон}$ с шагом dx . Интервал и шаг задать таким образом, чтобы проверить все ветви программы. Таблицу снабдить заголовком и шапкой.

Вариант 2

Провести интегрирование заданной функции методом прямоугольников на интервале от $X_{нач}$ до $X_{кон}$ с шагом dx . Проверить работу программы на функциях `math.sin`, `math.sqrt`, а также с помощью лямбда выражения `lambda x: x**2`

Вариант 3

Провести интегрирование заданной функции методом трапеций на интервале от $X_{нач}$ до $X_{кон}$ с шагом dx . Проверить работу программы на функциях `math.sin`, `math.sqrt`, а также с помощью лямбда выражения `lambda x: x**2`

Вариант 4

Вычислить среднее арифметическое заданного списка чисел. Не использовать встроенные функции и методы.

Вариант 5.

Вычислить среднее геометрическое списка чисел. Не использовать встроенные функции и методы.

Вариант 6.

Реализовать игру: пользователь вводит число, а компьютер отвечает, является ли число четным или нечетным. Выполнение повторяется до тех пор, пока пользователь не введет число 0.

Вариант 7.

С помощью цикла найти число пробелов во введенной пользователем строке.

Вариант 8.

Пользователь располагает суммой в 100 долларов. За каждый ход он вводит положительную (поступления) или отрицательную (затраты) сумму. Необходимо корректировать баланс и прервать игру, если сумма на счете станет отрицательной.

Вариант 9.

Вывести введенный текст «лесенкой» по буквам

Вариант 10.

Вывести на экран числа от 1 до 100. При этом вместо чисел, кратных трем, программа должна выводить слово Fizz, а вместо чисел, кратных пяти — слово Buzz. Если число кратно пятнадцати, то программа должна выводить слово FizzBuzz.

Задание к лабораторной работе № 5 (ЛР5) (8 баллов)

«Одномерные массивы»

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Сумму отрицательных элементов.
2. Произведение элементов, расположенных между максимальным и минимальным элементами. Упорядочить элементы массива по возрастанию.

Вариант 2

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Сумму положительных элементов.
2. Произведение элементов, расположенных между максимальным по модулю и минимальным по модулю элементами. Упорядочить элементы массива по убыванию.

Вариант 3

В одномерном массиве, состоящем из n целочисленных элементов, вычислить:

1. Произведение элементов с четными номерами.
2. Сумму элементов, расположенных между первым и последним нулевыми элементами. Преобразовать массив таким образом, чтобы сначала располагались все положительные элементы, а потом - все отрицательные (элементы, равные нулю, считать положительными).

Вариант 4

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Сумму элементов с нечетными номерами.
2. Сумму элементов, расположенных между первым и последним отрицательными элементами. Сжать массив, удалив из него все элементы, модуль которых не превышает единицу. Освободившиеся в конце массива элементы заполнить нулями.

Вариант 5

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Максимальный элемент массива.
2. Сумму элементов, расположенных до последнего положительного элемента. Сжать массив, удалив из него все элементы, модуль которых находится в интервале $[a, b]$. Освободившиеся в конце массива элементы заполнить нулями.

Вариант 6

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Минимальный элемент массива.
2. Сумму элементов, расположенных между первым и последним положительными элементами. Преобразовать массив таким образом, чтобы сначала располагались все элементы, равные нулю, а потом — все остальные.

Вариант 7

В одномерном массиве, состоящем из n целочисленных элементов, вычислить:

1. Номер максимального элемента массива.
2. Произведение элементов массива, расположенных между первым и вторым нулевыми элементами. Преобразовать массив таким образом, чтобы в первой его половине располагались элементы, стоявшие в нечетных позициях, а во второй половине - элементы, стоявшие в четных позициях.

Вариант 8

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Номер минимального элемента.
2. Сумму элементов, расположенных между первым и вторым отрицательными элементами. Преобразовать массив таким образом, чтобы сначала располагались все элементы, модуль которых не превышает единицу, а потом - все остальные.

Вариант 9

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Максимальный по модулю элемент.
2. Сумму элементов, расположенных между первым и вторым положительными элементами. Преобразовать массив таким образом, чтобы элементы, равные нулю, располагались после всех остальных.

Вариант 10

В одномерном массиве, состоящем из n целочисленных элементов, вычислить:

1. Минимальный по модулю элемент.
2. Сумму модулей элементов, расположенных после первого элемента, равного нулю. Преобразовать массив таким образом, чтобы в первой его половине располагались элементы, стоявшие в четных позициях, а во второй половине - элементы, стоявшие в нечетных позициях.

Вариант 11

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Номер минимального по модулю элемента.
2. Сумму модулей элементов, расположенных после первого отрицательного элемента. Сжать массив, удалив из него все элементы, величина которых находится в интервале $[a, b]$. Освободившиеся в конце массива элементы заполнить нулями.

Вариант 12

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Номер максимального по модулю элемента.
2. Сумму элементов, расположенных после первого положительного элемента. Преобразовать массив таким образом, чтобы сначала располагались все элементы, целая часть которых лежит в интервале $[a, b]$, а потом — все остальные.

Вариант 13

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Количество элементов массива, лежащих в диапазоне от A до B .
2. Сумму элементов, расположенных после максимального элемента. Упорядочить элементы массива по убыванию модулей.

Вариант 14

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Количество элементов массива, равных нулю.
2. Сумму элементов, расположенных после минимального элемента. Упорядочить элементы массива по возрастанию модулей.

Вариант 15

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Количество элементов массива, больших C .
2. Произведение элементов, расположенных после максимального по модулю элемента. Преобразовать массив таким образом, чтобы сначала располагались все отрицательные элементы, а потом - все положительные (элементы, равные нулю, считать положительными).

Вариант 16

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Количество отрицательных элементов.
2. Сумму модулей элементов, расположенных после минимального по модулю элемента. Заменить все отрицательные элементы массива их квадратами и упорядочить элементы массива по возрастанию.

Вариант 17

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Количество положительных элементов.
2. Сумму элементов, расположенных после последнего элемента, равного нулю. Преобразовать массив таким образом, чтобы сначала располагались все элементы, целая часть которых не превышает единицу, а потом - все остальные.

Вариант 18

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Количество элементов массива, меньших S .
2. Сумму целых частей элементов массива, расположенных после последнего отрицательного элемента.

Преобразовать массив таким образом, чтобы сначала располагались все элементы, отличающиеся от максимального не более чем на 20 %, а потом - все остальные:

Вариант 19

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Произведение отрицательных элементов.
2. Сумму положительных элементов, расположенных до максимального элемента.

Изменить порядок следования элементов в массиве на обратный.

Вариант 20

В одномерном массиве, состоящем из n вещественных элементов, вычислить:

1. Произведение положительных элементов.
2. Сумму элементов, расположенных до минимального элемента. Упорядочить по возрастанию отдельно элементы, стоящие на четных местах, и элементы, стоящие на нечетных местах.

Задание к лабораторной работе № 6 (ЛР6) (8 баллов)

«Двумерные массивы и функции»

Размерности двумерных массивов следует запрашивать у пользователя. Все необходимые данные должны передаваться в функции в качестве параметров. Все переменные, используемые только внутри функции, должны быть описаны как локальные. Использование глобальных переменных в функциях не допускается. Обеспечить вывод, как исходного массива, так и массива, полученного в результате работы программы, там, где это возможно по условию задачи.

Пункты задания оформить в виде функций.

Вариант 1

Дана целочисленная прямоугольная матрица. Определить:

1. Количество строк, не содержащих ни одного нулевого элемента.
2. Максимальное значение из чисел, встречающихся в заданной матрице более одного раза.

Вариант 2

Дана целочисленная прямоугольная матрица.

1. Определить количество столбцов, не содержащих ни одного нулевого элемента.
2. Переставляя строки заданной матрицы, расположить их в соответствии с ростом характеристик.

ПРИМЕЧАНИЕ: Характеристикой строки целочисленной матрицы назовем сумму ее положительных четных элементов.

Вариант 3

Дана целочисленная прямоугольная матрица. Определить:

1. Количество столбцов, содержащих хотя бы один нулевой элемент.
2. Номер строки, в которой находится самая длинная серия одинаковых элементов.

Вариант 4

Дана целочисленная квадратная матрица. Определить:

1. Произведение элементов в тех строках, которые не содержат отрицательных элементов.
2. Максимум среди сумм элементов диагоналей, параллельных главной диагонали матрицы.

Вариант 5

Дана целочисленная квадратная матрица. Определить:

1. Сумму элементов в тех столбцах, которые не содержат отрицательных элементов.
2. Минимум среди сумм модулей элементов диагоналей, параллельных побочной диагонали матрицы.

Вариант 6

Дана целочисленная прямоугольная матрица. Определить:

1. Сумму элементов в тех строках, которые содержат хотя бы один отрицательный элемент.
2. Номера строк и столбцов всех седловых точек матрицы.

ПРИМЕЧАНИЕ: Матрица A имеет седловую точку a_{ij} , если a_{ij} является минимальным элементом в i -й строке и максимальным в j -м столбце.

Вариант 7

Для заданной матрицы размером 8×8 найти такие k , что элементы k -й строки матрицы совпадают с элементами k -ого столбца.

Найти сумму элементов в тех строках, которые содержат хотя бы один отрицательный элемент.

Вариант 8

Переставляя столбцы заданной матрицы, расположить их в соответствии с ростом характеристик.

Найти сумму элементов в тех столбцах, которые содержат хотя бы один отрицательный элемент.

ПРИМЕЧАНИЕ: Характеристикой столбца целочисленной матрицы назовем сумму модулей его отрицательных нечетных элементов.



Критерии оценки:

- оценка «отлично» выставляется студенту, если выполнены все задания лабораторной работы, составлен отчет по работе, согласно требованиям;
- оценка «хорошо» выставляется студенту, если выполнены почти все задания, за исключением отдельных пунктов, лабораторной работы, составлен отчет по работе, согласно требованиям;
- оценка «удовлетворительно» выставляется студенту, если выполнены больше половины заданий лабораторной работы, составлен отчет по работе, согласно

требованиям;

- оценка «неудовлетворительно» выставляется студенту, если выполнены меньше половины заданий лабораторной работы и не составлен отчет по работе, согласно требованиям;

Темы эссе (рефератов, докладов, сообщений)

1. Численное статистическое моделирование.
2. Многомерное шкалирование.
3. Непараметрические методы обработки данных
4. Разработка интеллектуальных информационных систем
5. Веб-сервисы анализа данных
6. Визуализация данных с использованием пакетов статистического анализа
7. Построение точечных оценок генеральных параметров
8. Группировка и визуализация данных
9. Корреляционный анализ порядковых переменных
10. Построение и проверка адекватности линейной модели парной регрессии.

Реферат оценивается следующим образом:

- соответствие содержания теме- 4 балла;
- глубина проработки материала, 3 балла;
- грамотность и полнота использования источников, 1 балл;
- соответствие оформления реферата требованиям, 2 балла;
- доклад, 5 баллов;
- умение вести дискуссию и ответы на вопросы, 5 баллов.

Максимальное количество баллов: 20.

Критерии оценки:

- оценка «отлично» выставляется студенту, если набрал 19-20 баллов;
- оценка «хорошо» выставляется студенту, если набрал 15-18 баллов;
- оценка «удовлетворительно» выставляется студенту, если набрал 10-14 баллов;
- оценка «неудовлетворительно» выставляется студенту, если набрал менее 10 баллов;

Вопросы к экзамену

1. Основы программирования на языке Python (refresher).
2. Введение в анализ данных на языке Python.
3. Математический аппарат (refresher).

4. Введение в модуль NumPy.
5. Основы работы с Pandas.
6. Разведывательный анализ данных
7. Математический аппарат для анализа данных: векторы, матрицы, функции и производные.
8. Особенности типов данных в NumPy. Работа с векторами и матрицами. 9. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).
10. Введение в модуль для работы с табличным представлением данных Pandas.
11. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным.
13. Визуализация данных. Представление результатов исследования
14. Введение в визуализацию данных. Нюансы визуализации данных и принципы человеческого восприятия.
15. Правила создания хороших визуализаций.
16. Работа с текстовыми данными. Сбор данных из открытых источников. 17. Предварительная обработка текстовых данных. Текстовый анализ.
17. Особенности разных видов графиков и их использования.
18. Обзор языка разметки markdown.
19. Экспорт данных и отчетов из Jupyter Notebook.
20. Создание интерактивных визуализаций и отчетов с помощью инструмента Plotly.
21. Анализ сетей
22. Введение в модуль NetworkX.
23. Подготовка данных для анализа сетей. Примеры визуализации сетей на примере данных из сети Twitter.
24. Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.
25. Регрессионные модели. Метод наименьших квадратов. Логистическая регрессия.
26. Библиотека sklearn.
27. Деревья решений. Случайный лес. Ансамбли моделей
28. Решение задач кластеризации. Меры расстояния.
29. Введение в ансамбли моделей. Стэкинг и бэггинг. Случайный лес. Бустинг.
30. Введение в нейронные сети.
31. Проблемы обучения нейронными сетями.
32. Библиотеки для работы с нейронными сетями - tensorflow и keras.

Критерии оценки:

- «зачтено» выставляется студенту, если изложение полученных знаний в устной форме полное, в системе, в соответствии с требованиями учебной программы; допускаются, отдельные несущественные ошибки, исправляемые учащимися после указания преподавателя на них;
- «не зачтено» выставляется студенту, если изложение учебного материала неполное, бессистемное, что препятствует усвоению последующей учебной информации; существенные ошибки, не исправляемые даже с помощью преподавателя.

