МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет математики и компьютерных наук

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

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 участника образовательных отношений ОПОП*

Махачкала, 2023

Рабочая программа дисциплины «Статистический анализ в Python» составлена в 2023 году в соответствии с требованиями ФГОС ВО-бакалавриат по направлению подготовки *01.03.05 Статистика* от 14.08.2020 № 1032.

Разработчики:
кафедра прикладной математики: Гаджиева Т.Ю., к.ф.-м.н., доц.

Программа одобрена:
на заседании кафедры прикладной математики
от «20» 01 2023 г., протокол № 5

Зав. кафедрой _____ Кадиев Р. И.

на заседании Методической комиссии факультета математики и компьютерных наук от «25» 01 2023 г., протокол № 4

Председатель _____ Ризаев М.К.

Рабочая программа дисциплины согласована с учебно-методическим управлением «28» февраля 2023 г.

Начальник УМУ _____ Гасангаджиева А.Г.
(подпись)

Аннотация рабочей программы дисциплины

Дисциплина «Статистический анализ в Python» входит *в часть, формируемую участниками образовательных отношений ОПОП* образовательной программы *бакалавриата* по направлению подготовки 01.03.05 - Статистика.

Дисциплина реализуется на факультете математики и компьютерных наук кафедрой прикладной математики.

Содержание дисциплины «Статистический анализ в Python» охватывает круг вопросов о представлении, сборе, обработке и анализе данных ан языке программирования Python.

Дисциплина нацелена на формирование следующих компетенций выпускника: профессиональных – ПК-5.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: *лекции, лабораторные занятия, самостоятельная работа.*

Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости: выполнение лабораторной работы и промежуточный контроль в форме экзамена.

Объем дисциплины 3 зачетные единицы (108 академических часах), в том числе по видам учебных занятий

Семестр	Учебные занятия							Форма промежуточной аттестации (зачет, дифференцированный зачет, экзамен
	в том числе							
	Контактная работа обучающихся с преподавателем						СРС, в том числе экзамен	
	Всего	из них						
		Лекции	Лабораторные занятия	Практические занятия	КСР	консультации		
6	108	16	16				76	экзамен

1. Цели освоения дисциплины

Цель курса "Статистический анализ в Python" – развитие навыков статистических данных с использованием алгоритмов, созданных на языке Python, представления о сборе, обработке и анализе данных в интерактивной среду Jupyter, введение в автоматизированные методы работы с данными – машинное обучение и нейронные сети.

2.Место дисциплины в структуре ОПОП бакалавриата

Дисциплина «Статистический анализ в Python» входит в *часть, формируемую участниками образовательных отношений ОПОП* образовательной программы *бакалавриата* по направлению подготовки 01.03.05 - Статистика.

Курс по дисциплине «Статистический анализ в Python» вводится после изучения дисциплин теория вероятностей и математическая статистика, статистика предприятия «Математический анализ», «Теория вероятностей и математическая статистика», «Информатика», так как для успешного усвоения этого курса студентам необходимы знания по указанным дисциплинам.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения и процедура освоения).

Код и наименование компетенции из ОПОП	**Код и наименование индикатора достижения компетенций** *(в соответствии с ОПОП*	**Планируемые результаты обучения**	**Процедура освоения**
ПК-5. Способен разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи	ПК-5.1. Знает разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	Знает: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или	Участие в коллективной разработке проектов

		прикладной задачи. Умеет: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Владеет: навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	
	ПК-5.2. Умеет разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	Знает: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Умеет: разрабатывать и реализовывать в	

		виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.. Владеет: навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	
	ПК-5.3. Владеет навыками разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи	Знает: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи. Умеет: разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной	

		теоретической или прикладной задачи.. Владеет: навыками разработки и реализации в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи.	

4. Объем, структура и содержание дисциплины.

4.1. Объем дисциплины составляет 3 зачетных единиц, 108 академических часов.

4.2. Структура дисциплины.

№	Раздел Дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Самостоятельная работа	Формы текущего контроля успеваемости *(по неделям семестра)* Форма промежуточной аттестации *(по семестрам)*
				Лекции	Практические	Лаборат. занятия	Контр. сам.раб		
МОДУЛЬ 1: Методы и алгоритмы аналитики статистических данных									
1	Основы программирования на языке Python (refresher).	6	1	2		2		5	Индивидуальный фронтальный опрос, выполнение лабораторной работы
2	Математический аппарат (refresher). Введение в модуль NumPy. Основы работы с Pandas.	6	2	2		2		5	

	Разведывательный анализ данных								
3	Визуализация данных. Представление результатов исследования..	6	3	2		2		5	
4	Работа с текстовыми данными. Сбор данных из открытых источников. Предварительная обработка текстовых данных. Текстовый анализ.			2		2		5	
	Итого по модулю 1:	6		8		8		20	
МОДУЛЬ 2: Статистический анализ данных в Python									
5	Анализ сетей	6	4-5	2		2		5	Индивидуальный фронтальный опрос, выполнение лабораторной работы
6	Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.			2		2		5	
7	Деревья решений. Случайный лес. Ансамбли моделей.	6	6-7	2		2		5	
8	Введение в нейронные сети.	6	8	2		2		5	
	Итого по модулю 2:	6		8		8		20	
МОДУЛЬ 3: Экзамен									
7	Подготовка к экзамену	6						36	
	Итого по модулю 3:	6						36	
	ИТОГО по дисциплине:	**6**		**16**		**16**		**76**	

4.3. Содержание дисциплины, структурированное по темам (разделам).

4.3.1. Содержание лекционных занятий по дисциплине

МОДУЛЬ 1. Методы и алгоритмы аналитики статистических данных

Тема 1. Основы программирования на языке Python (refresher).

Введение в анализ данных на языке Python. Почему Python становится стандартом для работы с большими данными. Прикладные задачи политологов, для решения которых подходит язык Python.

Тема 2. Математический аппарат (refresher). Введение в модуль NumPy. Основы работы с Pandas. Разведывательный анализ данных

Математический аппарат для анализа данных: векторы, матрицы, функции и производные.

Особенные типы данных в NumPy. Работа с векторами и матрицами.

Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).

Введение в модуль для работы с табличным представлением данных Pandas.

Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным.

Тема 3. Визуализация данных. Представление результатов исследования

Введение в визуализацию данных. Нюансы визуализации данных и принципы человеческого восприятия. Правила создания хороших визуализаций.

Тема 4. Работа с текстовыми данными. Сбор данных из открытых источников. Предварительная обработка текстовых данных. Текстовый анализ.

Обзор различных видов графиков (гистограмма, бар-чарт, секторная диаграмма, линейные графики, график рассеяния, тепловая карта и т.д.). Особенности разных видов графиков и их использования. Обзор языка разметки markdown. Экспорт данных и отчетов из Jupyter Notebook. Создание интерактивных визуализаций и отчетов с помощью инструмента Plotly.

Введение в анализ текста. Применение в политологии. Особенности подготовки данных.

МОДУЛЬ 2: Статистический анализ данных в Python

Тема 5. Анализ сетей

Введение в анализ сетей. Основные метрики и параметры сетей. Введение в модуль NetworkX. Подготовка данных для анализа сетей. Примеры визуализации сетей на примере данных из сети Twitter.

Тема 6. Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.
Введение в машинное обучение. Обучение с учителем и без учителя. Проблема переобучения. Регрессионные модели. Метод наименьших квадратов. Логистическая регрессия. Библиотека sklearn. Постановка гипотезы, преобразование данных и построение регрессионных моделей для задачи с пассажирами “Титаника”.
Тема 7. Деревья решений. Случайный лес. Ансамбли моделей
Решение задач кластеризации. Меры расстояния. Обзор алгоритмов кластеризации (иерархические алгоритмы, алгоритмы квадратичной ошибки, выделение связных компонент). Построчный разбор алгоритма кластеризации. Проверка его работы на синтетических данных (ирисы, случайные формы). Применение алгоритма кластеризации к набору данных о политических режимах.
Введение в ансамбли моделей. Стэкинг и бэггинг. Случайный лес. Бустинг.
Тема 8. Введение в нейронные сети.

Введение в нейронные сети. Перцептроны. Проблемы обучения нейронными сетей. Прикладное использование в политологии. Библиотеки для работы с нейронными сетями - tensorflow и keras. Пример сверточной сети на наборе данных MNIST (распознавание рукописного текста).

4.3.2. Содержание лабораторных занятий по дисциплине

МОДУЛЬ 1. Методы и алгоритмы аналитики статистических данных

Тема 1. Основы программирования на языке Python (refresher).
Основы программирования на языке Python: типы данных и методы работы с ними (переменные, листы, словари, кортежи).
Тема 2. Математический аппарат (refresher). Введение в модуль NumPy. Основы работы с Pandas. Разведывательный анализ данных
Особенные типы данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).
Введение в модуль для работы с табличным представлением данных Pandas. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным.
Тема 3. Визуализация данных. Представление результатов исследования
Работа с табличными данными в Pandas на примере данных Всемирного банка по продолжительности жизни и данных о публикационной активности

словацких экономистовТема 4. Работа с текстовыми данными. Сбор данных из открытых источников. Создание различных видов визуализаций на синтетических данных и тренировочных наборах данных. Создание визуализаций для разведывательного анализа данных для набора данных о пассажирах “Титаника”.

МОДУЛЬ 2: Статистический анализ данных в Python

Тема 5. Анализ сетей

Введение в модуль NetworkX. Подготовка данных для анализа сетей. Примеры визуализации сетей на примере данных из сети Twitter.

Тема 6. Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.

Библиотека sklearn. Постановка гипотезы, преобразование данных и построение регрессионных моделей для задачи с пассажирами “Титаника”

Тема 7. Деревья решений. Случайный лес. Ансамбли моделей

Разбор примеров. Построение ансамблей моделей на наборе данных “Титаник”.

Тема 8. Введение в нейронные сети.

Разбор примеров применения нейронных сетей в политологии. Разбор статьи Anastasopoulos et al. (2017) “Political image analysis with deep neural networks?”

5. **Образовательные технологии**

Лекции проводятся с использованием меловой доски и мела. Параллельно материал транслируется на экран с помощью мультимедийного проектора. Семинарские занятия проводятся с использованием мела и меловой доски. Для проведения лекционных занятий необходима аудитория, оснащенная мультимедиа-проектором, экраном, доской, ноутбуком (с программным обеспечение для демонстрации слайд-презентаций).

Для проведения семинарских занятий необходима аудитория на 25 человек, оснащена доской.

6. Учебно-методическое обеспечение самостоятельной работы студентов.

6.1. Виды и порядок выполнения самостоятельной работы

1. Изучение рекомендованной литературы.

2. Решение задач.

3. Выполнение лабораторных работ

4. Подготовка к экзамену.

№	Виды самостоятельной работы	Вид контроля	Учебно-методич. Обеспечения
1	Изучение рекомендованной литературы	Устный опрос по разделам дисциплины	См. разделы 6.2, 7.2, 8, 9 данного документа
2	Решение задач	Проверка домашнего задания	См. разделы 6.2, 7.2, 8, 9 данного документа
3	Выполнение лабораторных работ	Проверка лабораторной работы	См. разделы 6.2, 7.2, 8, 9 данного документа
4	Подготовка к экзамену	Устный опрос, либо компьютерное тестирование	См. разделы 6.2, 7.2, 8, 9 данного документа

Текущий контроль: выполнение лабораторных работ, проверка домашнего задания.

Промежуточная аттестация: проверка лабораторных работ.

Текущий контроль успеваемости осуществляется непрерывно, на протяжении всего курса. Прежде всего, это устный опрос по ходу практических занятий, выполняемый для оперативной активизации внимания студентов и оценки их уровня усвоения тем. Результаты устного опроса учитываются при выборе индивидуальных задач для решения. Каждую неделю осуществляется проверка выполнения заданий.

Промежуточный контроль проводится в форме контрольной работы, в которых содержатся практические задачи и теоретические вопросы.

Итоговый контроль проводится либо в виде устного экзамена, либо в форме тестирования.

Оценка «отлично» ставится за уверенное владение материалом курса.

Оценка «хорошо» ставится при полном выполнении требований к прохождению курса и умении ориентироваться в изученном материале.

Оценка «удовлетворительно» ставится при достаточном выполнении требований к прохождению курса и владении конкретными знаниями по программе курса.

Оценка «неудовлетворительно» ставится, если требования к прохождению курса не выполнены и студент не может показать владение материалом.

6.2. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.

Пакет заданий для самостоятельной работы выдается в начале семестра, определяются предельные сроки их выполнения и сдачи.

Название раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения	Кол-во часов	Литература
МОДУЛЬ 1: Методы и алгоритмы аналитики статистических данных			
Основы программирования на языке Python (refresher).	Введение в анализ данных на языке Python. Почему Python становится стандартом для работы с большими данными. Прикладные задачи политологов, для решения которых подходит язык Python.	5	Основная: 2, 3 Дополнительная: 1, 2, 3
Математический аппарат (refresher). Введение в модуль NumPy. Основы работы с Pandas. Разведывательный анализ данных	Математический аппарат для анализа данных: векторы, матрицы, функции и производные. Особенные типы данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических	5	Основная: 2, 3 Дополнительная: 1, 2, 3

	метрик с помощью NumPy (среднее, медиана, мода, дисперсия). Введение в модуль для работы с табличным представлением данных Pandas. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным.		
Визуализация данных. Представление результатов исследования..	Введение в визуализацию данных. Нюансы визуализации данных и принципы человеческого восприятия. Правила создания хороших визуализаций.	5	Основная: 1, 2, 3, Дополнительная: 1, 2, 3
Работа с текстовыми данными. Сбор данных из открытых источников. Предварительная обработка текстовых данных. Текстовый анализ.	Обзор различных видов графиков (гистограмма, бар-чарт, секторная диаграмма, линейные графики, график рассеяния, тепловая карта и т.д.). Особенности разных видов графиков и их использования. Обзор языка разметки markdown. Экспорт данных и отчетов из Jupyter Notebook. Создание интерактивных визуализаций и отчетов с помощью инструмента	5	Основная: 2, 3 Дополнительная: 1, 2, 3

	Plotly. Введение в анализ текста. Применение в политологии. Особенности подготовки данных.		
МОДУЛЬ 2: Статистический анализ данных в Python			
Анализ сетей	Введение в анализ сетей. Основные метрики и параметры сетей. Введение в модуль NetworkX. Подготовка данных для анализа сетей. Примеры визуализации сетей на примере данных из сети Twitter.	5	Основная: 1, 2, 3, Дополнительная: 1, 2, 3
Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.	Задачи классификации и линейные модели. Введение в машинное обучение. Обучение с учителем и без учителя. Проблема переобучения. Регрессионные модели. Метод наименьших квадратов. Логистическая регрессия. Библиотека sklearn. Постановка гипотезы, преобразование данных и построение регрессионных моделей для задачи с пассажирами “Титаника”.	5	Основная: 2, 3 Дополнительная: 1, 2, 3
Деревья решений. Случайный лес. Ансамбли моделей.	Случайный лес. Ансамбли моделей Решение задач кластеризации. Меры расстояния. Обзор алгоритмов	5	Основная: 1, 2, 3, Дополнительная: 1, 2, 3

	кластеризации (иерархические алгоритмы, алгоритмы квадратичной ошибки, выделение связных компонент). Построчный разбор алгоритма кластеризации. Проверка его работы на синтетических данных (ирисы, случайные формы). Применение алгоритма кластеризации к набору данных о политических режимах. Введение в ансамбли моделей. Стэкинг и бэггинг. Случайный лес. Бустинг.		
Введение в нейронные сети.	Введение в нейронные сети. Перцептроны. Проблемы обучения нейронными сетей. Прикладное использование в политологии. Библиотеки для работы с нейронными сетями - tensorflow и keras. Пример сверточной сети на наборе данных MNIST (распознавание рукописного текста).	5	Основная: 1, 2, 3, Дополнительная: 1, 2, 3

7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.

7.1 Типовые контрольные задания

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. Особенные типы 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 сетей.
32. Библиотеки для работы с нейронными сетями - tensorflow и keras.

Примерные лабораторные задания:

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

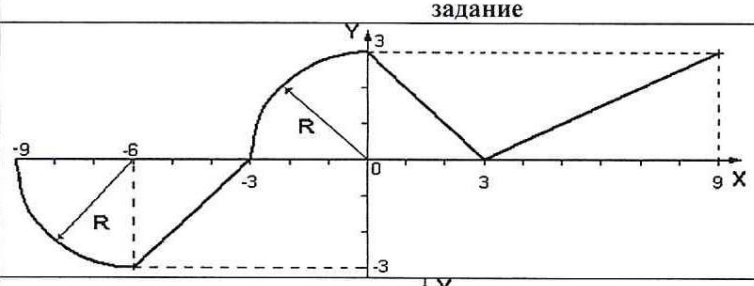
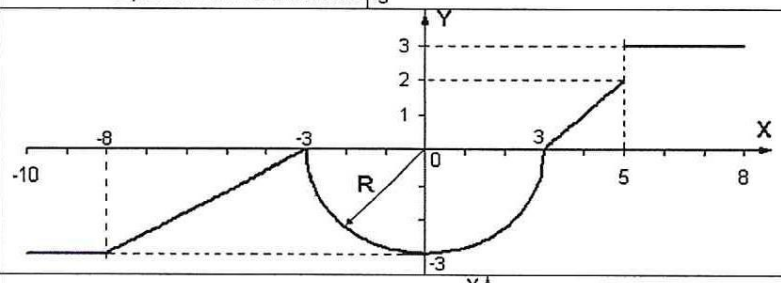
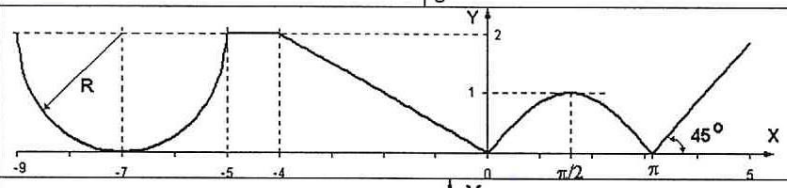
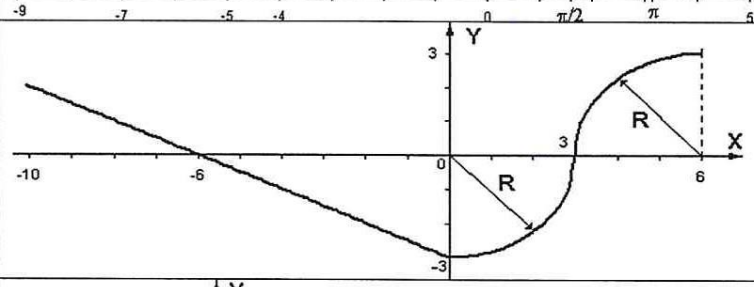
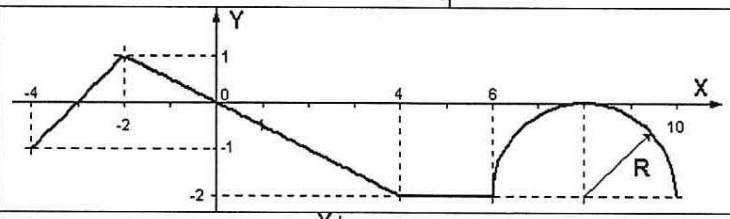
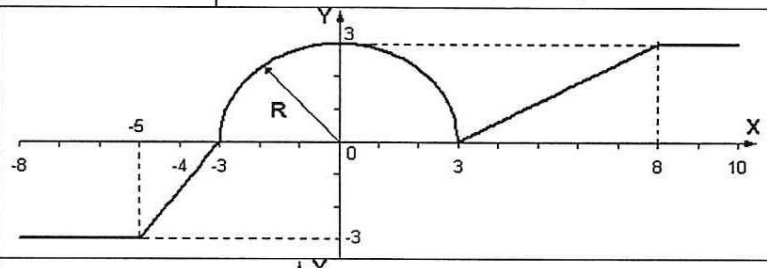
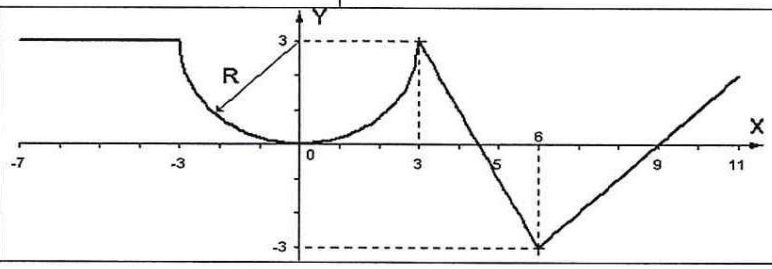
Описание алгоритма

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.

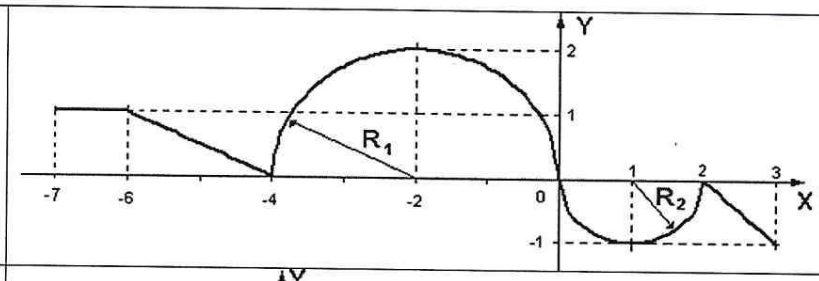
1. Ввести значения аи x, преобразовать к типу float.
2. Вычислить выражение1.
3. Вывести результатвычисления.
4. Вычислить значение переменной tmp;
5. Вычислить выражение2.
6. Вывести результатвычисления.

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 черезимеющиеся.

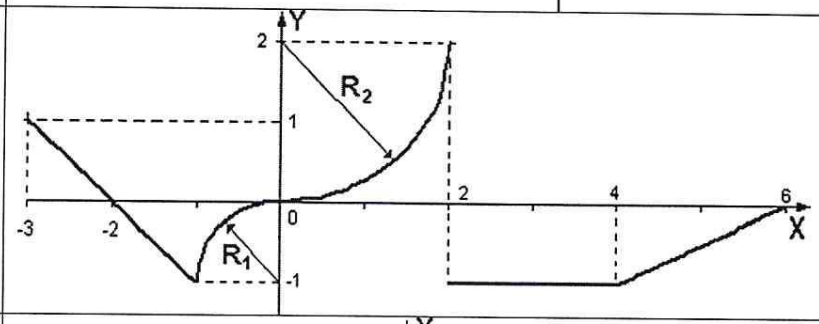
Таблица 1. Варианты заданий Лабораторной работы 2

вариант	задание
1	
2	
3	
4	
5	
6	
7	

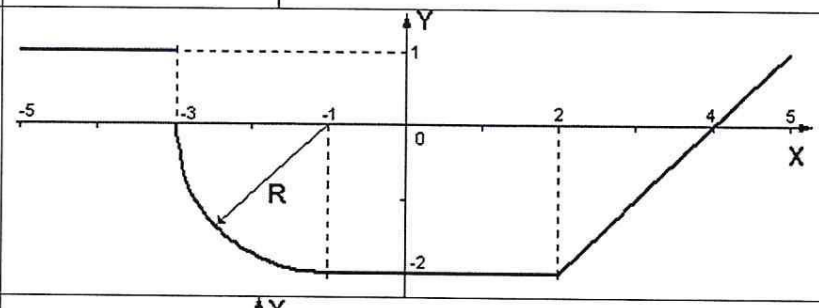
14



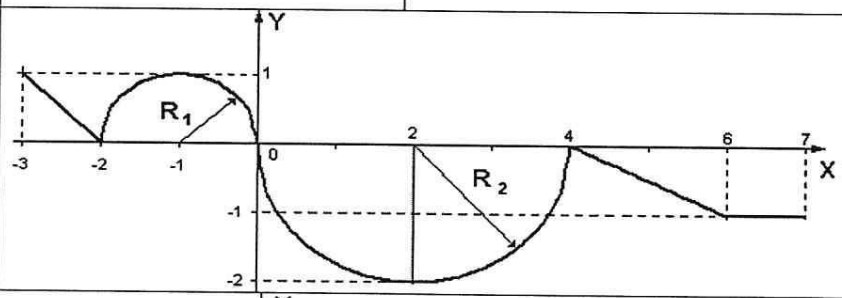
15



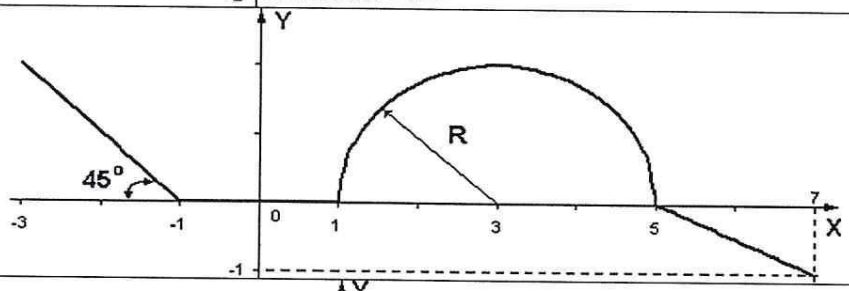
16



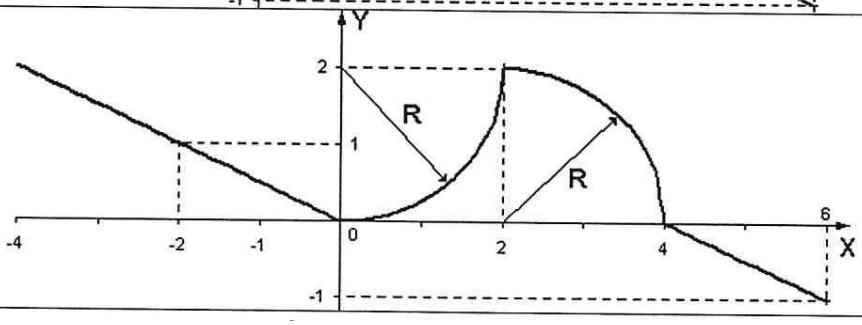
17

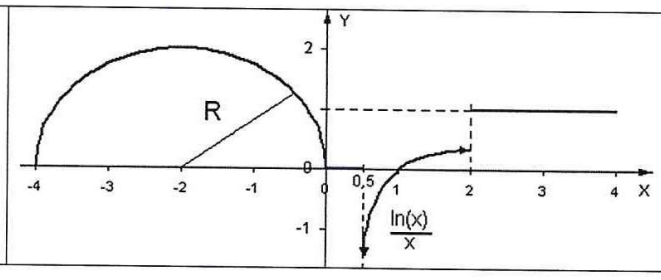


18



19



20	

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. Высчитать сумму чисел.
Вариант 2
Импортировать модуль NumPy. Создать одномерный массив случайных чисел. Высчитать 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-examsИмпортироватьмодуль 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 Хнач до Хкон 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 Хнач до Хкон с шагом dx. Проверить работу программы на функциях math.sin, math.sqrt, а также с помощью лямбда выражения lambdax:x**2

Вариант 3

Провести интегрирование заданной функции методом трапеций на интервале от Хнач до Хкон с шагом dx. Проверить работу программы на функциях math.sin, math.sqrt, а также с помощью лямбда выражения lambdax: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 повозрастанию.

Вариант 2

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

1. Сумму положительныхэлементов.
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 четныминомерами.
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, считатьположительными).

Вариант 4

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

1. Сумму элементов с нечетныминомерами.
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 элементмассива.
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 элементмассива.
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 п целочисленных элементов, вычислить:

1. Номер максимального элементамассива.
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 четныхпозициях.

Вариант 8

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

1. Номер минимальногоэлемента.
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 - всеостальные.

Вариант 9

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

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

Вариант 10

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

1. Минимальный по модулюэлемент.
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 модулюэлемента.
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 модулюэлемента.
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 убываниюмодулей.

Вариант 14

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

1. Количество элементов массива, равныхнулю.
2. Сумму элементов, расположенных после минимального элемента. Упорядочить элементы массива по возрастаниюмодулей.

Вариант 15

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

1. Количество элементов массива, большихC.
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 отрицательныхэлементов.
2. Сумму модулей элементов, расположенных после минимального по модулюэлемента.

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 положительныхэлементов.
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, меньшихС.
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 отрицательныхэлементов.
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 положительныхэлементов.
2. Сумму элементов, расположенных до минимального элемента. Упорядочить по возрастанию отдельно элементы, стоящие на четных местах, и элементы, стоящие на нечетныхместах.

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 нулевогоэлемента.
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 ростомхарактеристик.

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 нулевойэлемент.
2. Номер строки, в которой находится самая длинная серия одинаковых элементов.

Вариант4

Дана целочисленная квадратная матрица. Определить:

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

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 диагоналиматрицы.

Вариант 6

Дана целочисленная прямоугольная матрица. Определить:

1. Сумму элементов в тех строках, которые содержат хотя бы один отрицательныйэлемент.
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 x 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 всоответствии с ростомхарактеристик.

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

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

Вариант 9

Элемент матрицы называется локальным минимумом, если он строго меньше всех имеющихся у него соседей. Понятие соседей дано в варианте 9.

Подсчитать количество локальных минимумов заданной матрицы размером 10 x10.

Найти сумму модулей элементов, расположенных выше главной диагонали.

Вариант 10

Коэффициенты системы линейных уравнений заданы в виде прямоугольной матрицы. С помощью допустимых преобразований привести систему к треугольному виду.

Найти количество строк, среднее арифметическое элементов которых меньше заданной величины.

Вариант 11

Осуществить циклический сдвиг элементов прямоугольной матрицы на n элементов вправо или вниз (в зависимости от введенного режима), n может быть больше количества элементов в строке или столбце.

Вариант 12

Уплотнить заданную матрицу, удаляя из нее строкии столбцы, заполненныеенулями.

Найти номер первой из строк, содержащих хотя бы один положительныйэлемент.

Вариант 13

Дана целочисленная прямоугольная матрица. Определить номер первого из столбцов, содержащих хотя бы один нулевойэлемент.

Переставляя строки заданной матрицы, расположить их в соответствии с убыванием характеристик.

ПРИМЕЧАНИЕ: Характеристикой строки целочисленной матрицы назовем сумму ее отрицательных четныхэлементов.

Вариант 14

Упорядочить строки целочисленной прямоугольной матрицы по возрастанию количества одинаковых элементов в каждойстроке.

Найти номер первого из столбцов, не содержащих ни одного отрицательного элемента.

Вариант 15

Путем перестановки элементов квадратной вещественной матрицы добиться того, чтобы ее максимальный элемент находился в левом верхнем углу (1,1), следующий по величине – в позиции (2, 2), следующий по величине – в позиции (3, 3) и т. д., заполнив, таким образом, всю главную диагональ.

Найти номер первой из строк, не содержащих ни одного положительного элемента.

Вариант 16

Дана целочисленная прямоугольная матрица. Определить:

1. Количество строк, содержащих хотя бы один нулевойэлемент.
2. Номер столбца, в котором находится самая длинная серия одинаковых элементов.

Вариант 17

Дана целочисленная прямоугольная матрица. Определить:

1. Количество отрицательных элементов в тех строках, которые содержат хотя бы один нулевой 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 18

Написать программу, которая меняет местами столбцы квадратной матрицы, содержащие наибольший и наименьший элементы и вычисляет сумму элементов главной диагонали.

Вариант 19

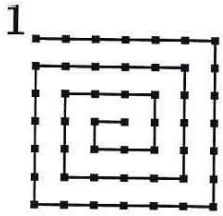
Дана целочисленная квадратная матрица. Определить:

1. Сумму элементов в тех строках, которые не содержат отрицательных элементов.
2. Минимум среди сумм элементов диагоналей, параллельных главной диагоналиматрицы.

Вариант 20

Напишите программу, формирующую квадратную матрицу, элементы которой являются натуральными числами, расположенными в порядке возрастания от 1 до n^2 (n – порядок матрицы) согласно схеме, приведённой на рисунке.

Вычислить сумму элементов, расположенных на главной диагонали полученной матрицы.



7.3. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Общий результат выводится как интегральная оценка, складывающая из текущего контроля – 30 % и промежуточного контроля – 70 %.

Текущий контроль по дисциплине включает:

- посещение занятий - 30 баллов,

- участие на практических занятиях - 35 баллов,

- выполнение домашних (аудиторных) контрольных работ - 35 баллов.

Промежуточный контроль по дисциплине включает:

- устный опрос - 40 баллов,

- письменная контрольная работа - 60 баллов,

Студенту выставляется:
- отлично, если интегральная оценка составляет 86 - 100 баллов;
- хорошо, если интегральная оценка составляет 66 - 85 баллов;
- удовлетворительно, если интегральная оценка составляет 51 - 65 баллов;
- неудовлетворительно, если интегральная оценка составляет 0 - 50 баллов.

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.

а) основная литература:

1. Буйначев, С. К. Основы программирования на языке Python : учебное пособие / С. К. Буйначев, Н. Ю. Боклаг ; под редакцией Ю. В. Песин. — Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2014. — 92 с. — ISBN 978-5-7996-1198-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.iprbookshop.ru/66183.html
2.Специализация “Python for Everybody”
https://www.coursera.org/specializations/python.
3. Эдвард Тафти (1983). Визуальное представление больших объемов информации.

б) дополнительная литература:

1. Васильев, А. Н. Python на примерах : практический курс по программированию / А. Н. Васильев. — 2-е изд. — Санкт-Петербург : Наука и Техника, 2017. — 432 с. — ISBN 978-5-94387-741-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.iprbookshop.ru/73043.html
2. Шелудько, В. М. Основы программирования на языке высокого уровня Python : учебное пособие / В. М. Шелудько. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2017. — 146 с. — ISBN 978-5-9275-2649-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.iprbookshop.ru/87461.html
3. Wes McKinney (2011). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.

1. Федеральный портал российское образование http://edu.ru;
2. Электронные каталоги Научной библиотеки Даггосуниверситета http://elib.dgu.ru/?q=node/256;
3. Образовательные ресурсы сети Интернет http://catalog.iot.ru/index.php;
4. Электронная библиотека http://elib.kuzstu.ru.

10. Методические указания для обучающихся по освоению дисциплины.

Дисциплина «Статистический анализ в Python» содержит внутри 3 модуля. Эти модули имеют определенную логическую завершенность по отношению к установленным целям и результатам обучения. Именно при изучении этих модулей должны развиваться компетенции ПК-5 применительно к теории массового обслуживания.

При изучении дисциплины рекомендуется рейтинговая технология обучения, которая позволяет реализовать комплексную систему оценивания учебных достижений студентов. Текущие оценки усредняются на протяжении семестра при изучении модулей. Комплексность означает учет всех форм учебной и творческой работы студента в течение семестра.

Рейтинг направлен на повышение ритмичности и эффективности самостоятельной работы студентов. Он основывается на широком использовании тестов и заинтересованности каждого студента в получении более высокой оценки знаний по дисциплине.

Принципы рейтинга: непрерывный контроль (в идеале на каждом из аудиторных занятий) и получение более высокой оценки за работу, выполненную в срок. При проведении практических занятий необходимо предусматривать широкое использование активных и интерактивных форм (компьютерных симуляций, деловых и ролевых игр).

Рейтинг включает в себя два вида контроля: текущий, промежуточный и итоговый по дисциплине.

Текущий контроль (ТК) - основная часть рейтинговой системы, основанная на беглом опросе раз в неделю или в две недели. Формы: тестовые оценки в ходе практических занятий, оценки за выполнение индивидуальных заданий и лабораторных работ. Важнейшей формой ТК, позволяющей опросить всех студентов на одном занятии являются короткие тесты из 2-3 тестовых заданий.

Основная цель ТК: своевременная оценка успеваемости студентов, побуждающая их работать равномерно, исключая малые загрузки или перегрузки в течение семестра.

Лекционные занятия желательно проводить в режиме презентаций с демонстрацией применения основных методов анализа и синтеза. Это существенно улучшает динамику лекций.

Целесообразно обеспечивать студентов на 1-2 лекции вперед раздаточным материалом в электронном виде (сложные схемы, графики, аналитические исследования и опорный конспект). Основное время лекции лучше тратить на подробные аналитические комментарии и особенности применения рассматриваемого материала в профессиональной деятельности студента.

Практические занятия следует проводить, используя профессиональные программы.

Промежуточный контроль (ПК) - это проверка знаний студентов по разделу программы. Формы: контрольная работа. Тестирование проводится в компьютерных классах в часы самостоятельной работы студентов по заранее составленному расписанию.

Цель ПК: побудить студентов отчитаться за усвоение раздела дисциплины накопительным образом, т.е. сначала за первый, затем за второй, затем за третий разделы каждого семестра.

Итоговый контроль по дисциплине (ИКД) - это проверка уровня учебных достижений студентов по всей дисциплине за семестр. Формы контроля: зачет.

Цель итогового контроля: проверка базовых знаний по дисциплине, полученных при изучении всех модулей семестра.

ИКД является выходным контролем по дисциплине, после которого можно рассчитывать на то, что процесс обучения по дисциплине завершен и в дальнейшем студент может сам при необходимости совершенствовать свои знания.

Распределение объемов различного вида контролей можно проиллюстрировать следующими цифрами на примере семестра: текущий контроль – 15 условных баллов; промежуточный контроль - 35 условных баллов; итоговый контроль - 50 условных баллов. Вся дисциплина оценивается в 100 условных баллов, если вся дисциплина оценивается цифрой, отличной от 100 баллов, то под условным баллом следует понимать процент от максимального числа баллов.

При этом действует следующая система перевода рейтинговых (условных) баллов в обычную шкалу оценок: “Отлично” (5) - 86−100 условных баллов; “Хорошо” (4) - 66−85 условных баллов; “Удовлетворительно” (3) - 51−65 условных баллов; “Неудовлетворительно” (2) - < 51 условных баллов.

Приведенные цифры говорят о том, что на любой стадии обучение студента можно считать удовлетворительным, если он набирает не менее 51 условных баллов. Так, например, набрав в ходе ТК и ПК 51 баллов, студент гарантирует себе оценку “удовлетворительно”.

Примеры оценочных средств (тестовых заданий) для текущего промежуточного и выходного контроля успеваемости по дисциплине:

Первый уровень сложности тестовых заданий (ТЗ) соответствует удовлетворительному владению предметом. Он представляет минимум базовых знаний, необходимых для дальнейшего обучения в университете и включает в себя знания - копии ключевых понятий и формул. Проверке этого уровня посвящены простейшие тестовые задания с нормой трудности в 1 балл.

Второй уровень ТЗ соответствует хорошим знаниям и предполагает глубокое понимание понятий и формул, умения их преобразовывать и интерпретировать.

Проверке второго уровня посвящены тестовые задания повышенной трудности, с нормой трудности в 2 балла.

Третий уровень ТЗ соответствует отличным знаниям и предполагает навыки по использованию ключевых понятий и формул в стандартных, а иногда

и в не стандартных ситуациях. Проверке третьего уровня посвящены наиболее трудные задания, с нормой трудности в 3 балла.

Задания каждого уровня снабжены соответствующими обозначениями. Это позволяет адаптивно строить усвоение программы дисциплины, когда каждый студент по мере усвоения курса на более низком уровне будет пробовать себя на более высоком уровне.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.

Для успешного освоения дисциплины, обучающий использует также кроме указанных выше в п. 8 программные обеспечения и интернет ресурсов: пакеты прикладных программ Python, Statistica.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

Учебные аудитории для проведения лекционных, семинарских и лабораторных занятий, компьютерные классы факультета и ИВЦ ДГУ. В университете имеется пакет необходимого лицензионного программного обеспечения.

При кафедре прикладной математики имеется студенческая научно-исследовательская лаборатория «Математическое моделирование».