Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Институт информационные системы и анализ данных

Центр программной инженерии

**ОТЧЁТ**

к лабораторной работе №2 по дисциплине:

|  |
| --- |
| «Методы анализа данных» |
| Методы корреляционного и регрессионного анализа |
| наименование темы |

Выполнилстудент ИСТб-21-1 Д.И. Морозов

номер группы подпись И. О. Фамилия

дата

Проверил Доцент Е.А. Осипова

Должность подпись И. О. Фамилия

дата

Иркутск – 2023 г.

# 1 Постановка задачи

А) Выбрать среду программирования для языка Python. Использовать набор данных согласно номеру варианта (таблица 2.1).

Б) Загрузить набор данных, подготовить данные для анализа. С использованием Python выполнить корреляционный и регрессионный анализ данных. При этом необходимо:

* получить значения основных показателей описательной статистики и построить диаграммы рассеяния для признаков, измеренных в количественной шкале;
* построить корреляционную матрицу и график «тепловая карта»;
* проверить гипотезы о значимости коэффициентов корреляции;
* для заданных пар количественных признаков построить уравнения линейной регрессии, построить диаграммы остатков, оценить дисперсию остатков;
* для заданных пар количественных признаков найти коэффициенты уравнения полиномиальной регрессии, построить его график и построить диаграммы остатков, оценить дисперсию остатков, сравнить диаграммы остатков с диаграммами остатков для линейной регрессии;
* для заданных пар количественных признаков оценить индекс корреляции, сравнить со значением коэффициента корреляции;
* выбрать переменную отклика и переменные факторы, построить для них модель множественной регрессии, объяснить полученные значения коэффициентов множественной регрессии.

Г) Выполнить анализ полученных на каждом этапе результатов и оформить отчет по лабораторной работе.

Вариант 1:

boston.csv (<https://www.kaggle.com/fedesoriano/the-boston-houseprice-data>)

Фактор – DIS;

Отклик – NOX.

# 2 Ход работы

# 2.1 Определение датасета

В качестве датасета был выбран - Вариант 1:

boston.csv (<https://www.kaggle.com/fedesoriano/the-boston-houseprice-data>)

Фактор – DIS;

Отклик – NOX.

# 2.2 Исходные данные и основные показатели описательной статистики

Датасет содержит 506 строку данных, и 14 признаков, измеренных в количественной шкале

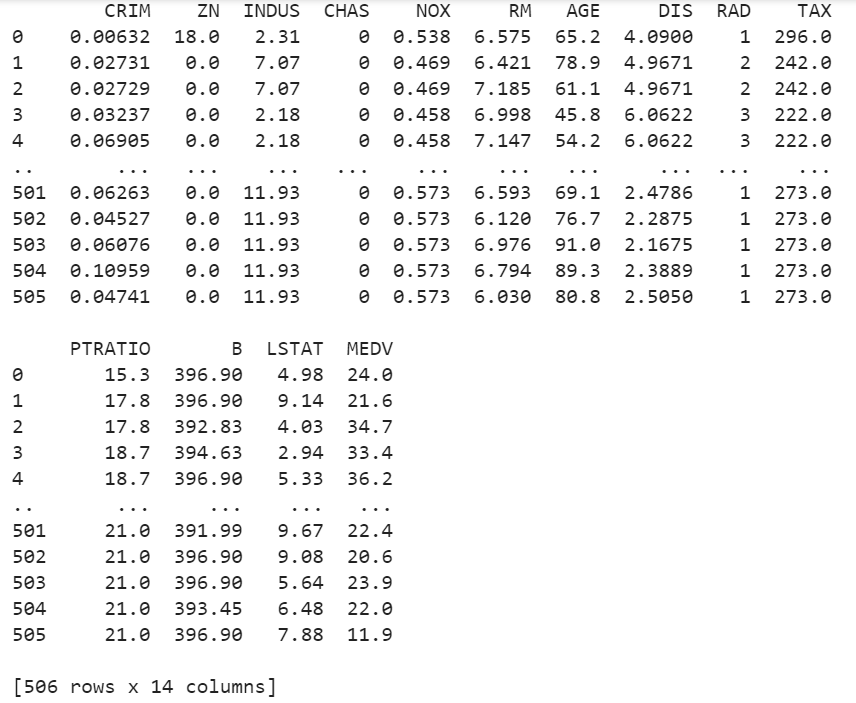


Рисунок 2.2.1 – Исходные данные.

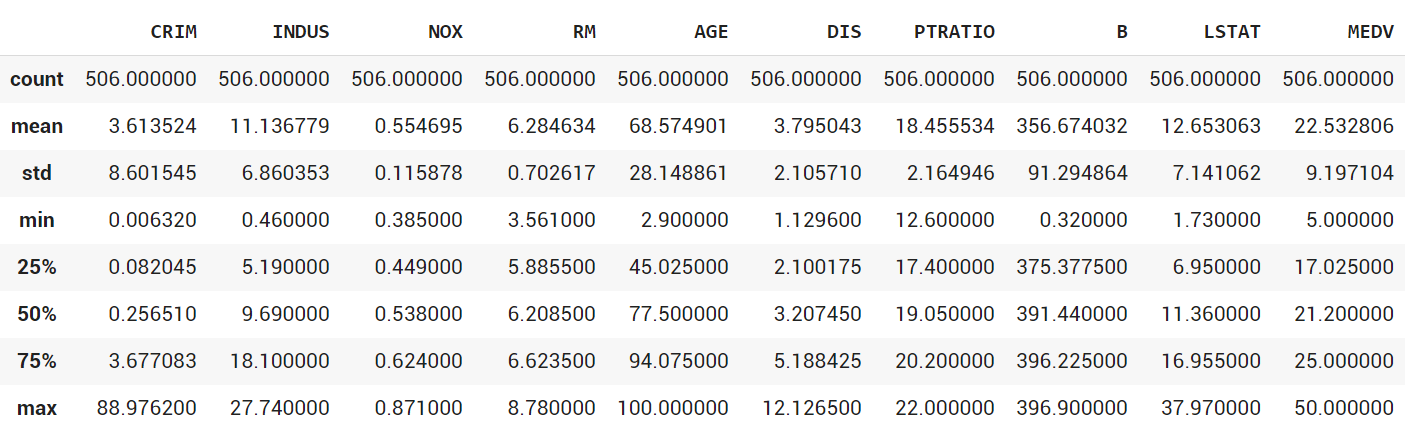


Рисунок 2.2.2 – Основные показатели описательной статистики.

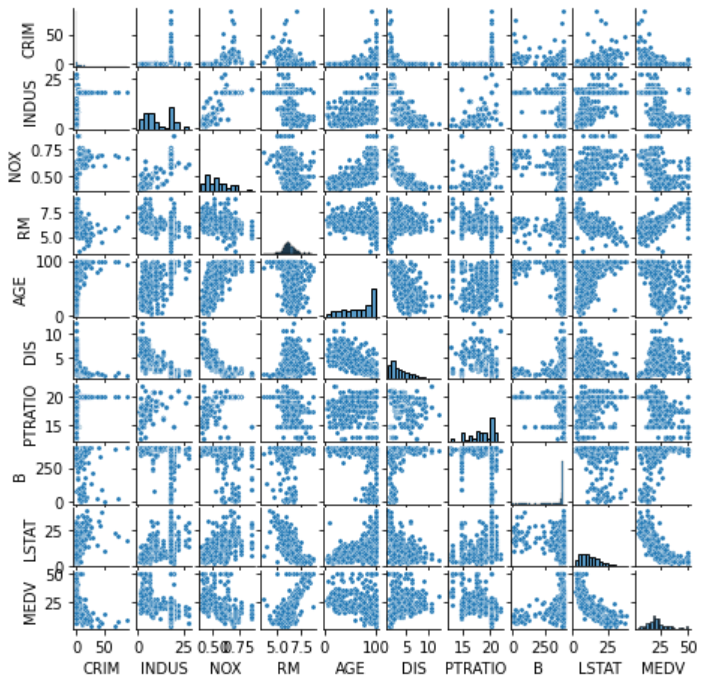


Рисунок 2.2.3 – Диаграммы рассеяния для признаков, измеренных в количественной шкале.

# 1.3 Корреляционная матрица и график «тепловая карта»

Построим корреляционную матрицу и график «тепловая карта»:

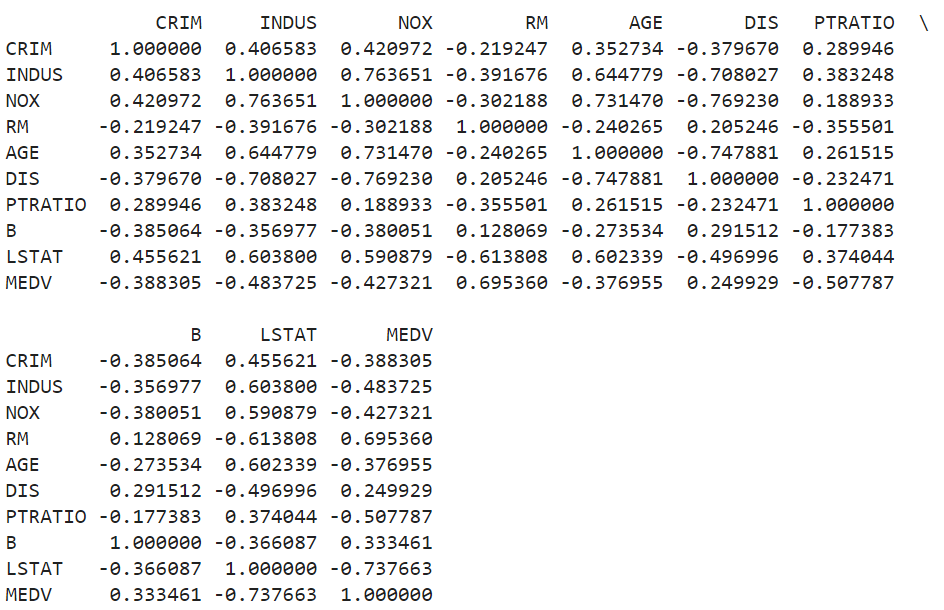


Рисунок 2.3.1 – Корреляционная матрица.

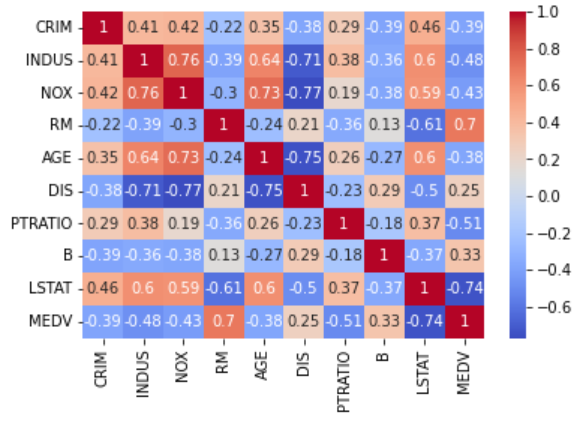
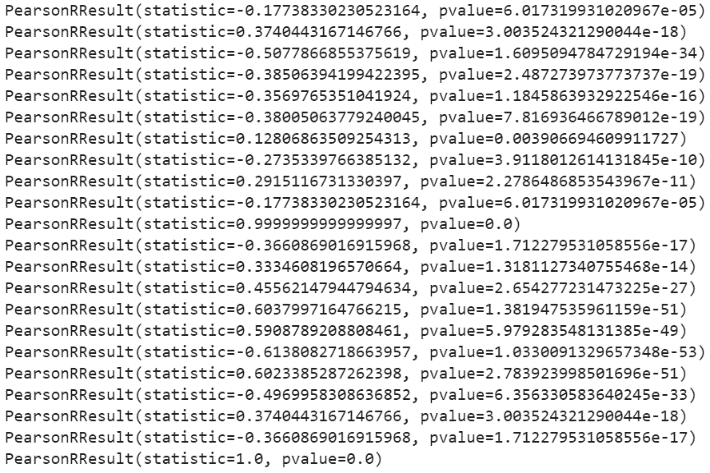
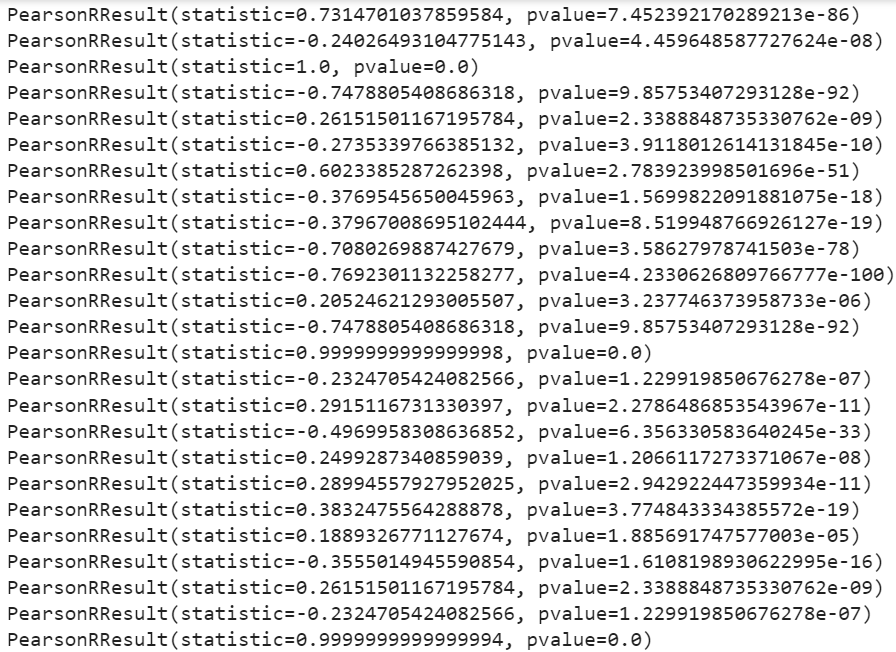


Рисунок 2.3.2 – График «тепловая карта».

# 2.4 Проверка гипотез о значимости коэффициентов корреляции

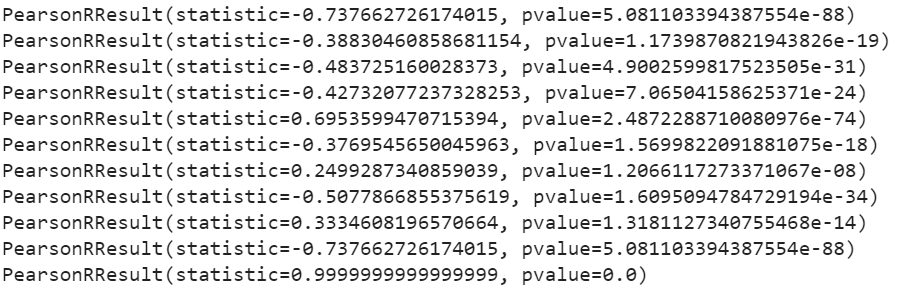


Рисунок 2.4.1 - Проверка гипотез о значимости коэффициентов корреляции

**2.5 Проверка значимости коэффициента корреляции по t-статистике**

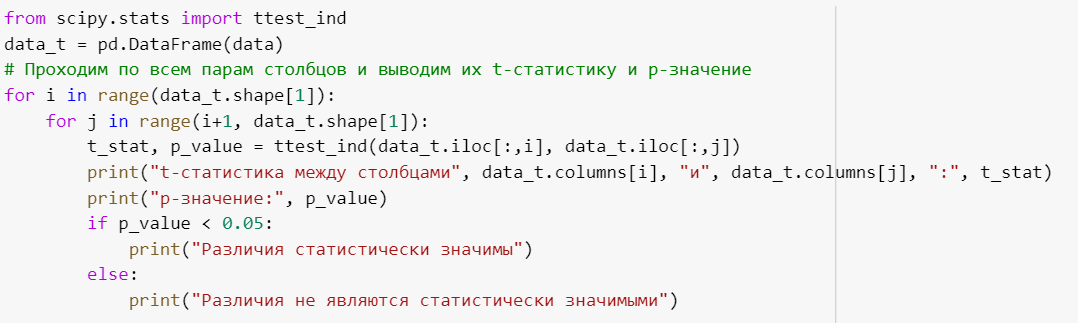


Рисунок 2.5.1 – Код для проверки по t-статистике

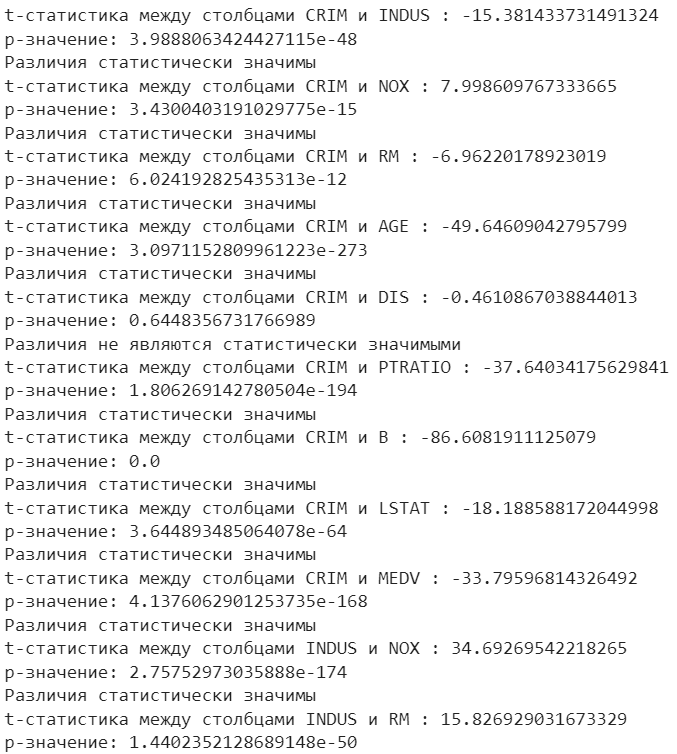
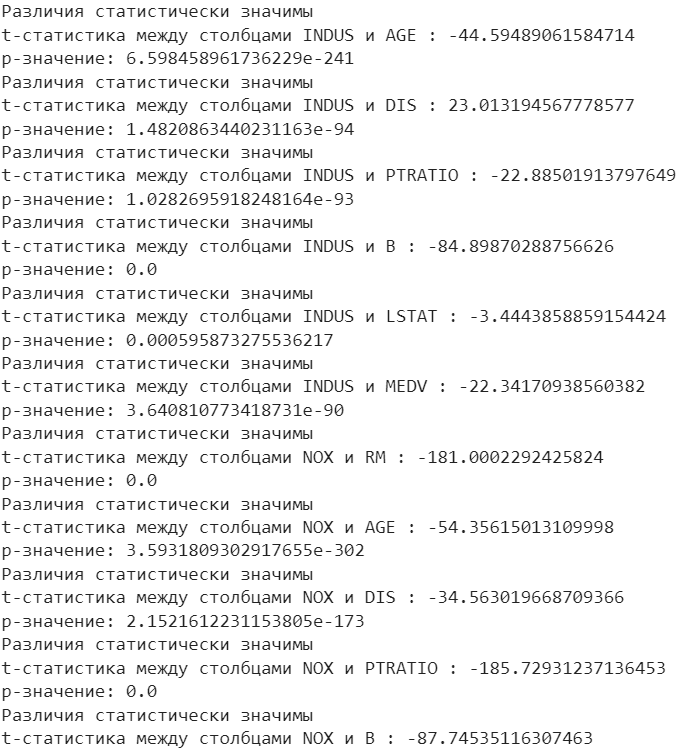


Рисунок 2.5.2 –Проверка значимости коэффициента корреляции по т статистике и

**2.6 Нахождение коэффициентов регрессии для пары «DIS» и «NOX»**

Для пары количественных признаков «DIS» - «NOX» построим уравнение линейной регрессии и диаграмму остатков регрессии, оценим дисперсию остатков:

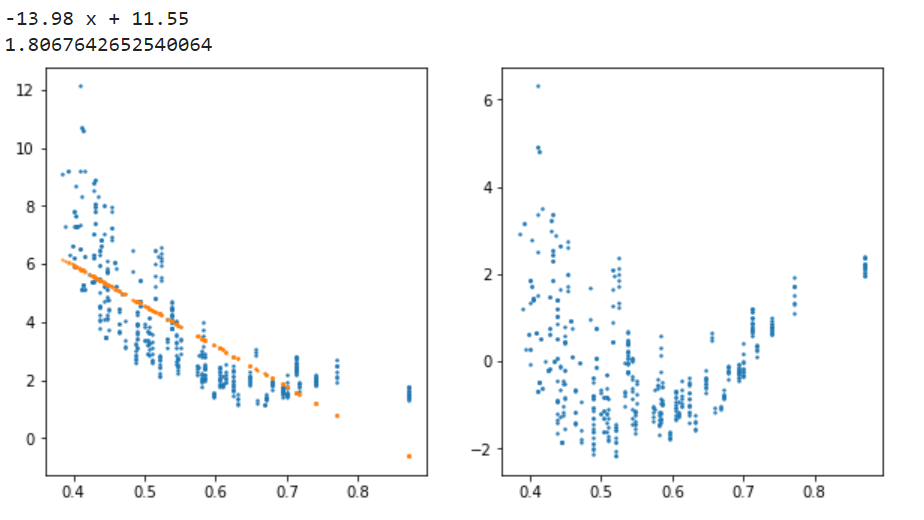


Рисунок 2.6.1 - Уравнение линейной регрессии и диаграммa остатков регрессии

Для пары количественных признаков «DIS» - «NOX» найдем коэффициенты уравнения полиномиальной регрессии, построим его график и диаграмму остатков регрессии, оценим дисперсию остатков:

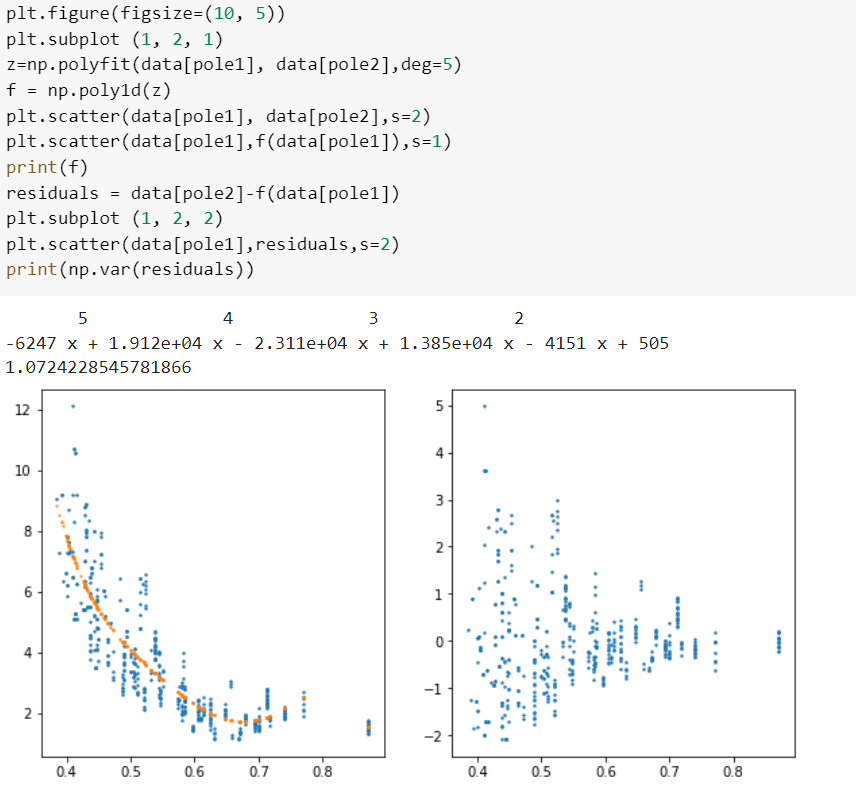


Рисунок 2.6.2 - Уравнение нелинейной регрессии и диаграммa остатков регрессии

**2.7 Нахождение коэффициентов регрессии для пары «DIS» и «NOX»**

Для пары количественных признаков «DIS» - «NOX» найдем значение индекса корреляции Пирсона и оценим индекс корреляции:

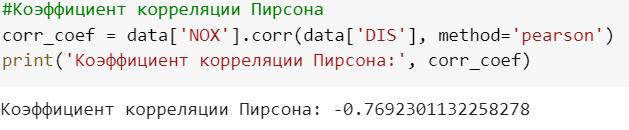


Рисунок 2.7.1 - Индекс корреляции Пирсона

**2.8 Множественная регрессия**

Исследуем зависимость длины головы опоссума от размера черепа и размера живота. Для этого построим модель множественной регрессии:

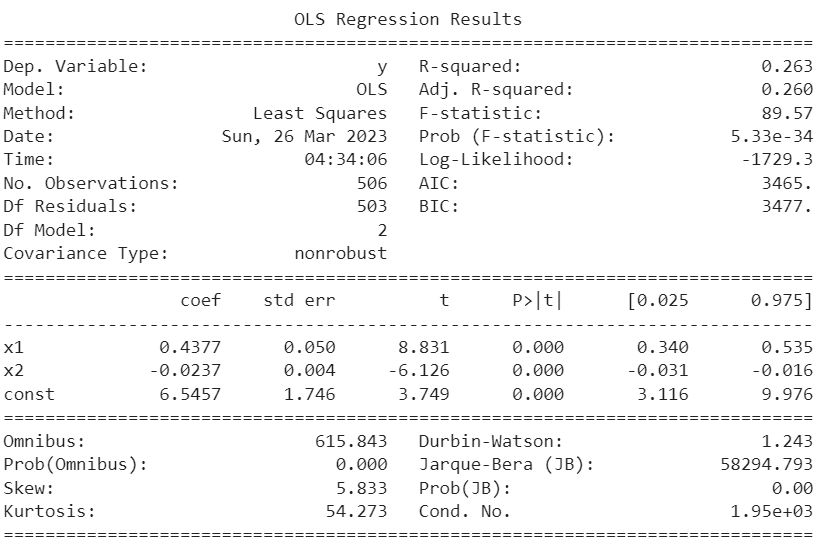


Рисунок 2.8.1 - Модель множественной регрессии

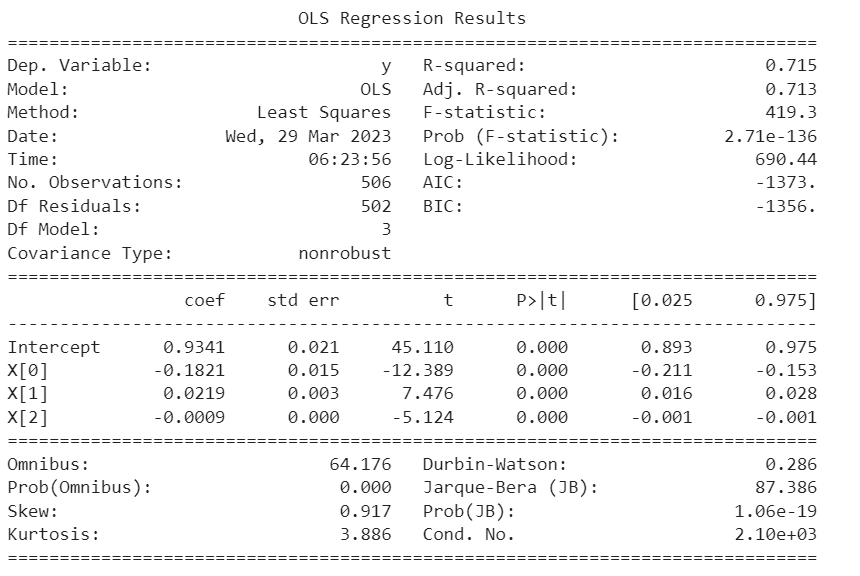


Рисунок 2.8.2 – Модель полиноминальной регрессии

3 Результаты работы

В данной работе проводится исследование методов поиска корреляционных зависимостей между признаками. Для этого были разработаны навыки построения матрицы корреляции и "тепловой карты", которые позволяют визуализировать зависимости между признаками. Кроме того, были построены линейные и нелинейные уравнения регрессии для признаков варианта 1, что позволяет оценить степень взаимосвязи между ними и предсказать значения одного признака на основе другого. Таким образом, проведенное исследование позволяет оценить взаимосвязь между признаками и предсказать значения одного признака на основе другого. Это может быть полезно в различных областях, таких как экономика, медицина, социология и другие.

# Список использованных источников

1 Петров А.В., Бучнев О.С. Лабораторный практикум — Иркутск: ФГБОУ ВО ИРНИТУ, 2022. — 114 с.