|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | | МИНОБРНАУКИ РОССИИ | | | | Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА** – **Российский технологический университет»**  **РТУ МИРЭА** | | | | |
| Институт информационных технологий (ИТ) | |
| Кафедра прикладной математики | |

|  |  |
| --- | --- |
| **ОТЧЁТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ № 4** | |
| **по дисциплине «Технологии и инструментарий анализа больших данных»** | |
|  | |
| Выполнил студент группы ИКБО-20-21 | Фомичев Р.А. |
| Проверил ассистент кафедры ПМ ИИТ | Тетерин Н.Н. |

Определить два вектора, представляющие собой число автомобилей, припаркованных в течении 5 рабочих дней у бизнес-центра на уличной стоянке и в подземном гараже.

1.1. Найти и интерпретировать корреляцию между переменными «Улица» и «Гараж» (подсчитать корреляцию по Пирсону).

Код программы представлен на рисунке 1.

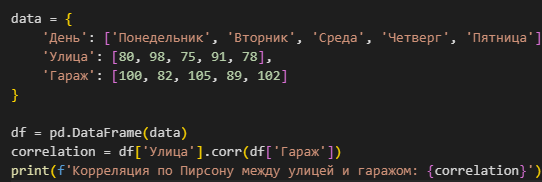


Рисунок 1 – Код программы

Вывод программы представлен на рисунке 2.

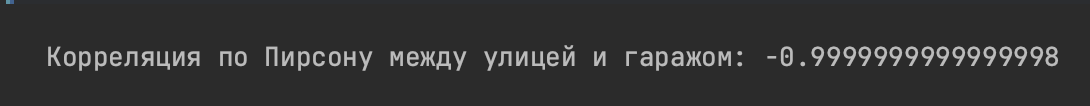


Рисунок 2 – Вывод программы

1.2. Построить диаграмму рассеяния для вышеупомянутых переменных.

Код программы представлен на рисунке 3.

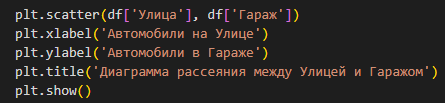


Рисунок 3 – Код программы

Вывод программы представлен на рисунке 4.

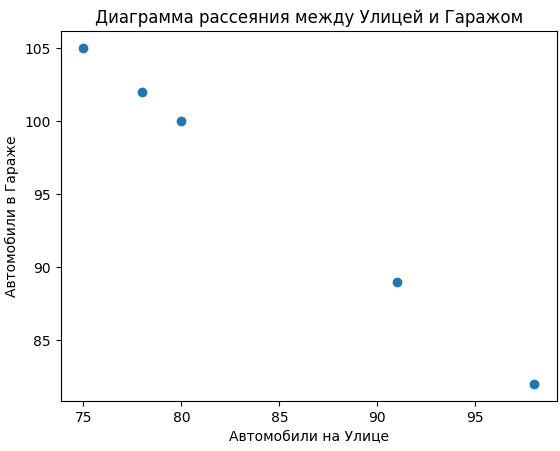


Рисунок 4 – Вывод программы

2. Найти и выгрузить данные. Вывести, провести предобработку и описать признаки.

Код программы представлен на рисунке 5.

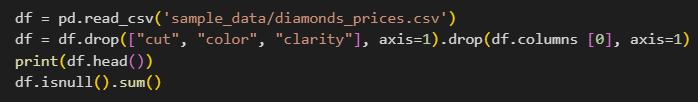


Рисунок 5 – Код программы

Вывод программы представлен на рисунке 6.

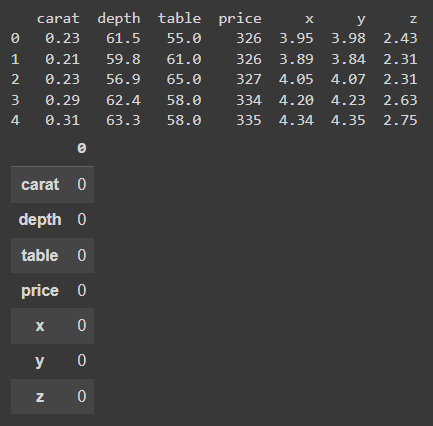


Рисунок 6 – Код программы

2.1. Построить корреляционную матрицу по одной целевой переменной. Определить наиболее коррелирующую переменную, продолжить с ней работу в следующем пункте.

Код программы представлен на рисунке 7.

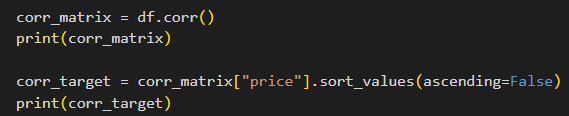


Рисунок 7 – Код программы

Вывод программы представлен на рисунке 8.

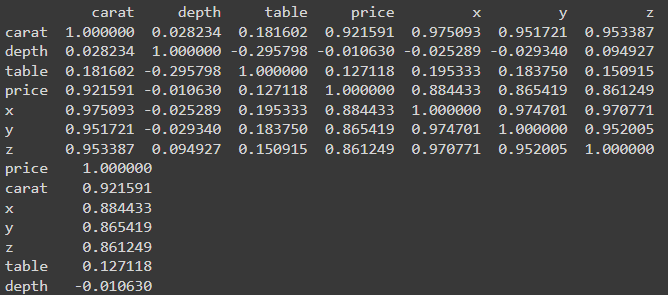


Рисунок 8 – Вывод программы

2.2. Реализовать регрессию вручную, отобразить наклон, сдвиг и MSE.

Код программы представлен на рисунке 9.

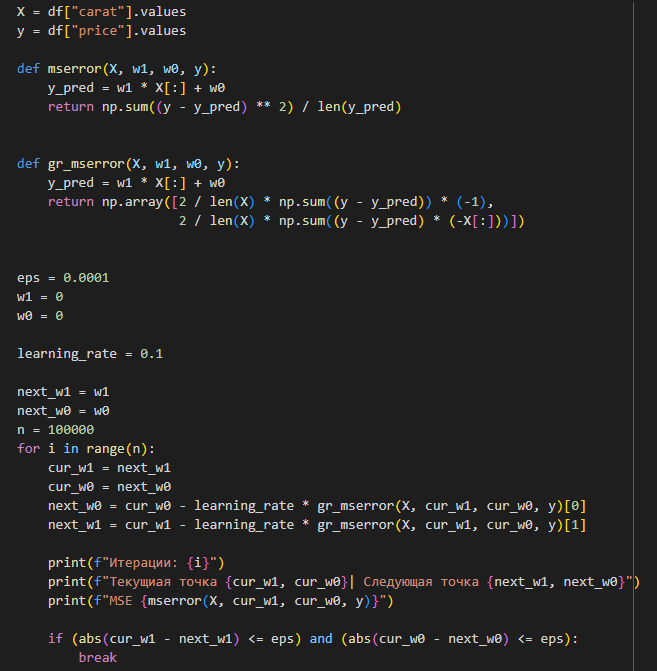


Рисунок 9 – Код программы

Вывод программы представлен на рисунке 10.

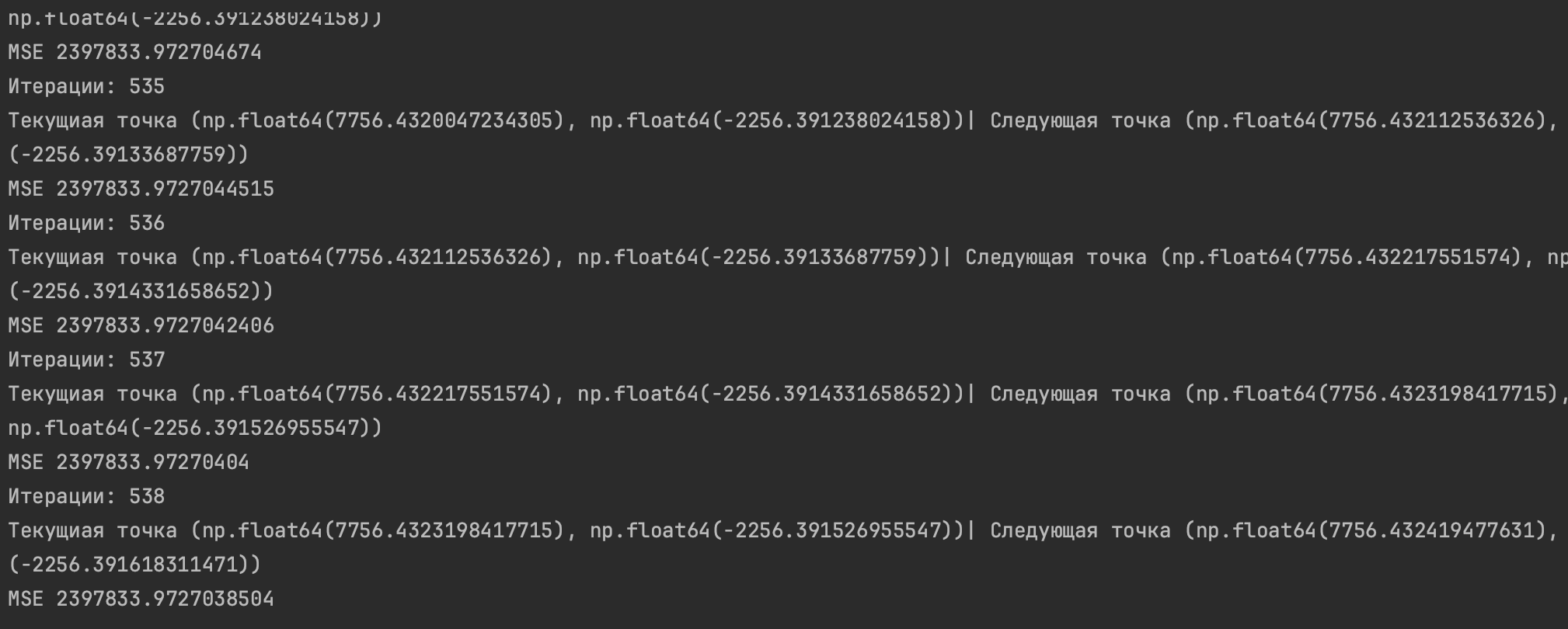


Рисунок 10 – Вывод программы

2.3. Визуализировать регрессию на графике.

Код программы представлен на рисунке 11.

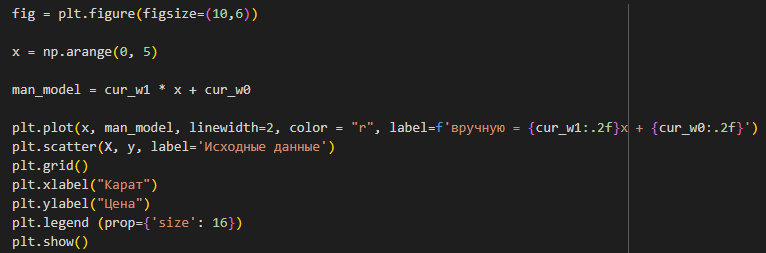


Рисунок 11 – Код программы

Вывод программы представлен на рисунке 12.

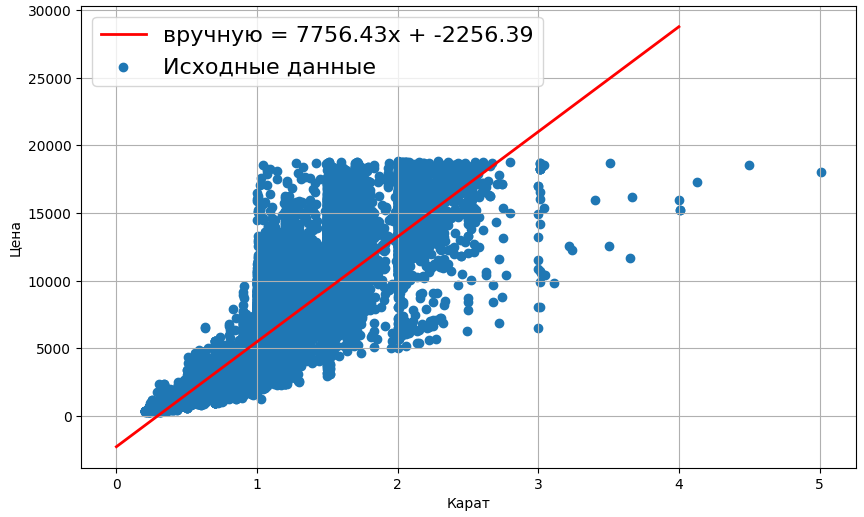


Рисунок 12 – Вывод программы

3. Загрузить данные: 'insurance.csv'. Вывести и провести предобработку. Вывести список уникальных регионов.

Код программы представлен на рисунке 13.

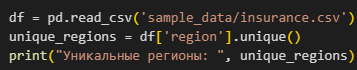


Рисунок 13 – Код программы

Вывод программы представлен на рисунке 14.



Рисунок 14 – Вывод программы

3.1. Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя первый способ, через библиотеку Scipy.

Код программы представлен на рисунке 15.

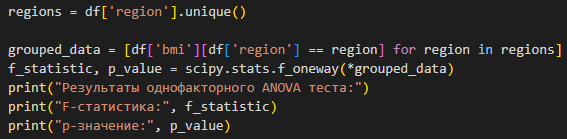


Рисунок 15 – Код программы

Вывод программы представлен на рисунке 16.



Рисунок 16 – Вывод программы

3.2. Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя второй способ, с помощью функции anova\_lm() из библиотеки statsmodels.

Код программы представлен на рисунке 17.

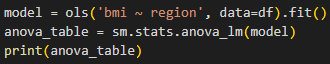


Рисунок 17 – Код программы

Вывод программы представлен на рисунке 18.

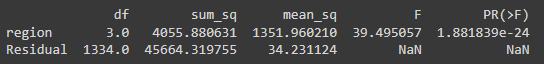


Рисунок 18 – Вывод программы

3.3. С помощью t критерия Стьюдента перебрать все пары. Определить поправку Бонферрони. Сделать выводы.

Код программы представлен на рисунке 19.

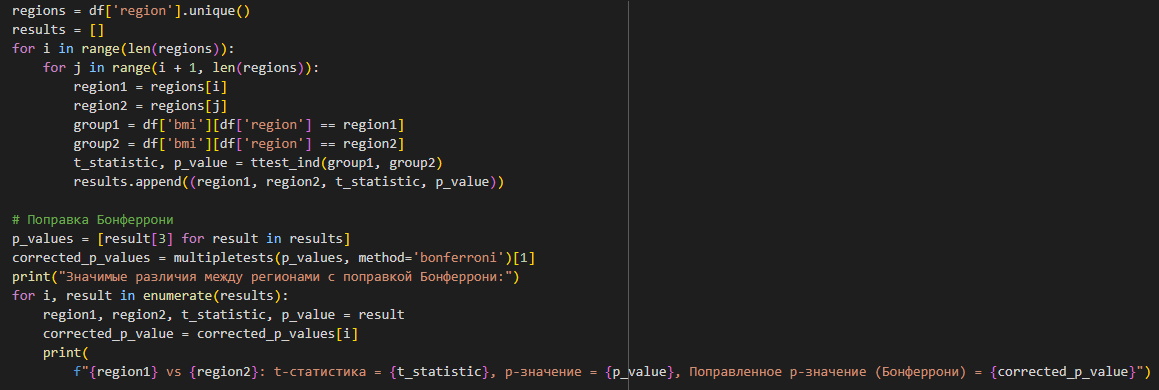


Рисунок 19 – Код программы

Вывод программы представлен на рисунке 20.

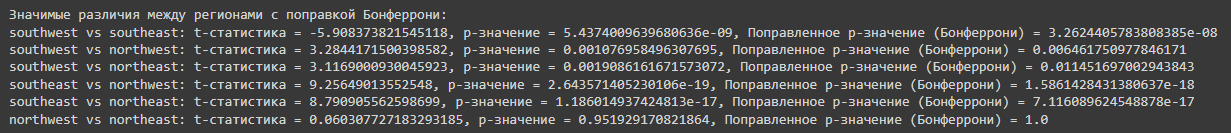


Рисунок 20 – Вывод программы

3.4. Выполнить пост-хок тесты Тьюки и построить график.

Код программы представлен на рисунке 21.

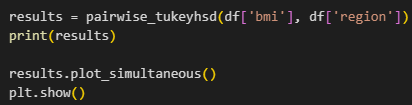


Рисунок 21 – Код программы

Вывод программы представлен на рисунках 22, 23.

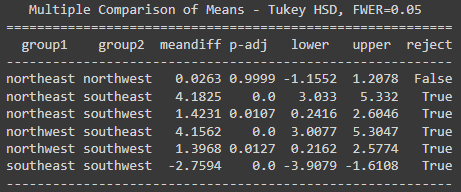


Рисунок 22 – Вывод программы

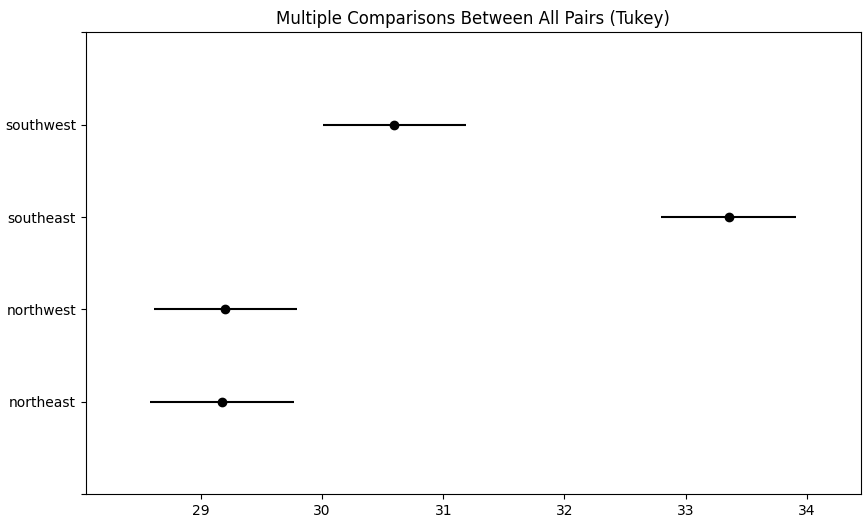


Рисунок 23 – Вывод программы

3.5. Выполнить двухфакторный ANOVA тест, чтобы проверить влияние региона и пола на индекс массы тела (BMI), используя функцию anova\_lm() из библиотеки statsmodels.

Код программы представлен на рисунке 24.

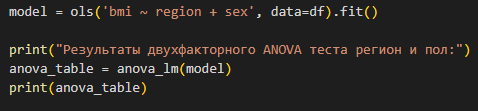


Рисунок 24 – Код программы

Вывод программы представлен на рисунке 25.

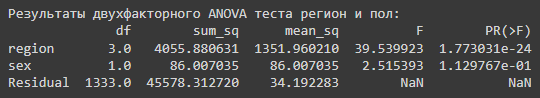


Рисунок 25 – Вывод программы

3.6. Выполнить пост-хок тесты Тьюки и построить график.

Код программы представлен на рисунке 26.

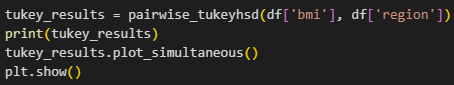


Рисунок 26 – Код программы

Вывод программы представлен на рисунках 27, 28.

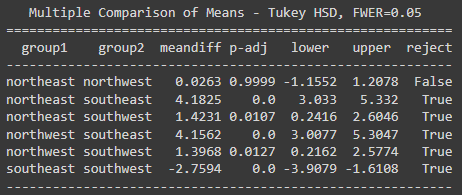


Рисунок 27 – Вывод программы

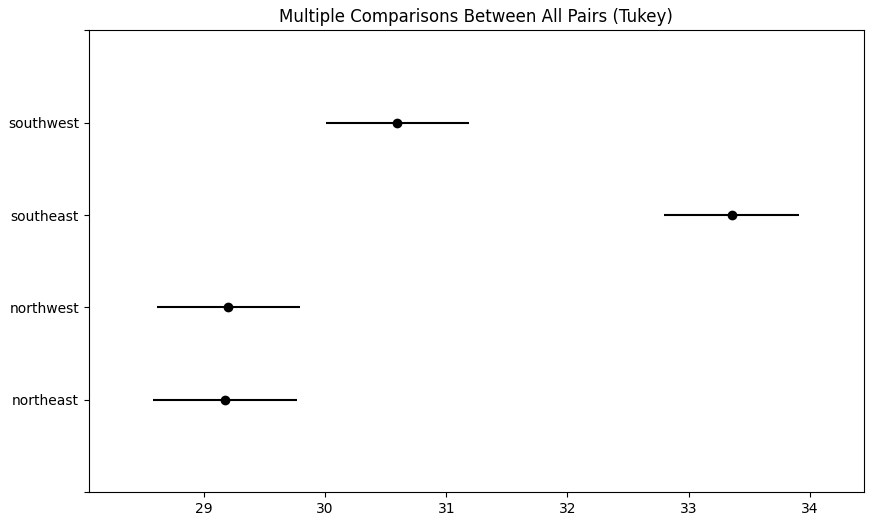


Рисунок 28 – Вывод программы