|  |
| --- |
|  |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования  **"МИРЭА - Российский технологический университет"**  **РТУ МИРЭА** |
| Институт информационных технологий (ИТ) |
| Кафедра прикладной математики (ПМ) |

|  |  |
| --- | --- |
| **ОТЧЕТ**  **ПО ПРАКТИЧЕСКОЙ РАБОТЕ №8** | |
| **по дисциплине** |  |
| **«Технологии и инструментарий анализа больших данных»** | |
| Выполнил студент группы ИКБО-20-19 | Московка А.А. |
| Принял семинарист кафедры Прикладной Математики | Десятников А.А. |

|  |  |  |
| --- | --- | --- |
| Практическая работы выполнены | « » 2022г. |  |
| «Зачтено» | « » 2022 г. |  |

Москва 2022

**Постановка задачи**

1. Загрузить данные: 'insurance.csv'. Вывести и провести предобработку. Вывести список уникальных регионов.
2. Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя первый способ, через библиотеку Scipy.
3. Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя второй способ, с помощью функции anova\_lm() из библиотеки statsmodels.
4. С помощью t критерия Стьюдента перебрать все пары. Определить поправку Бонферрони. Сделать выводы.
5. Выполнить пост-хок тесты Тьюки и построить график.
6. Выполнить двухфакторный ANOVA тест, чтобы проверить влияние региона и пола на индекс массы тела (BMI), используя функцию anova\_lm() из библиотеки statsmodels.
7. Выполнить пост-хок тесты Тьюки и построить график.
8. Оформить отчет о проделанной работе, написать выводы.

**Программный код с комментариями и результаты его выполнения**



Рисунок 1 – Скриншот программного кода (Часть 1)



Рисунок 2 – Скриншот программного кода (Часть 2)



Рисунок 3 – Скриншот программного кода (Часть 3)



Рисунок 5 – Скриншот программного кода (Часть 4)

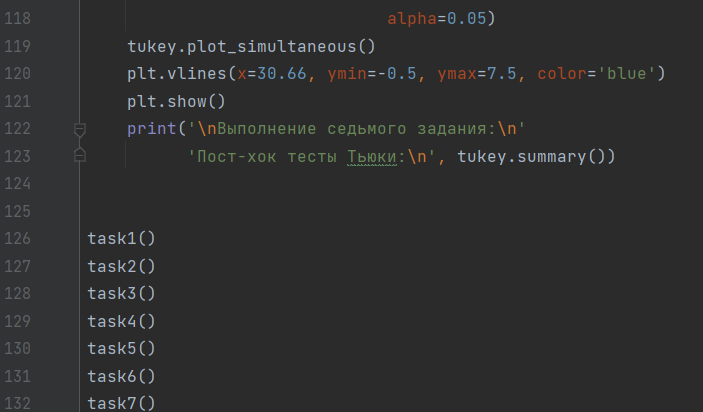


Рисунок 6 – Скриншот программного кода (Часть 5)

На следующих рисунках представлены скриншоты результатов выполнения программы со всеми заданиями с пояснениями, а также диаграммы (Рисунок 7-12).

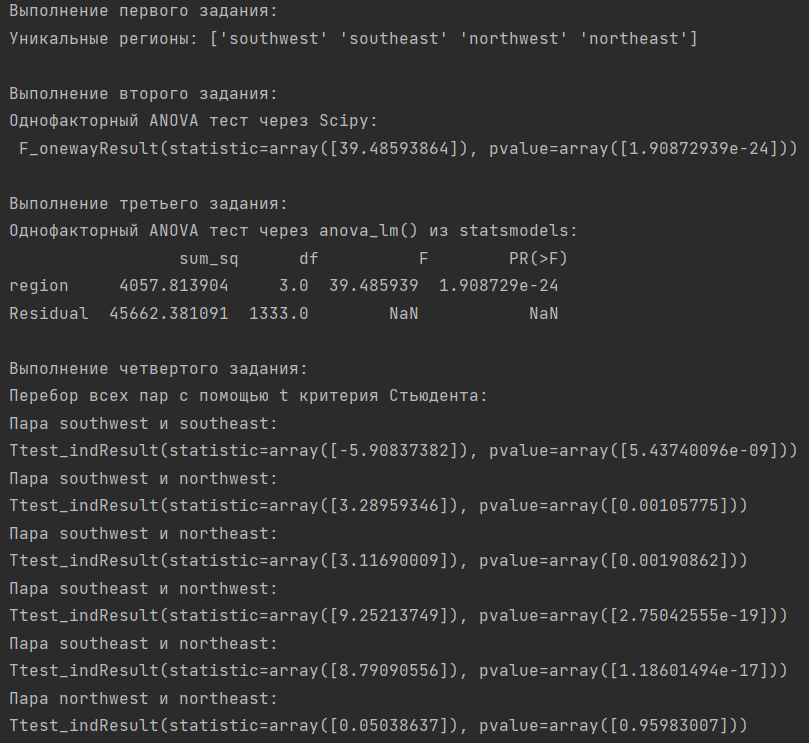


Рисунок 7 – Скриншот результата выполнения первого, второго третьего и четвертого заданий

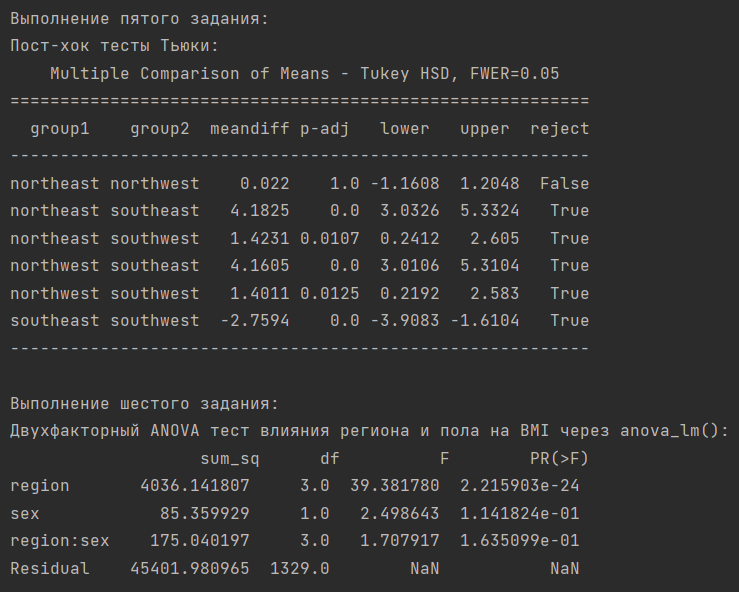


Рисунок 8 – Скриншот результата выполнения пятого и шестого заданий

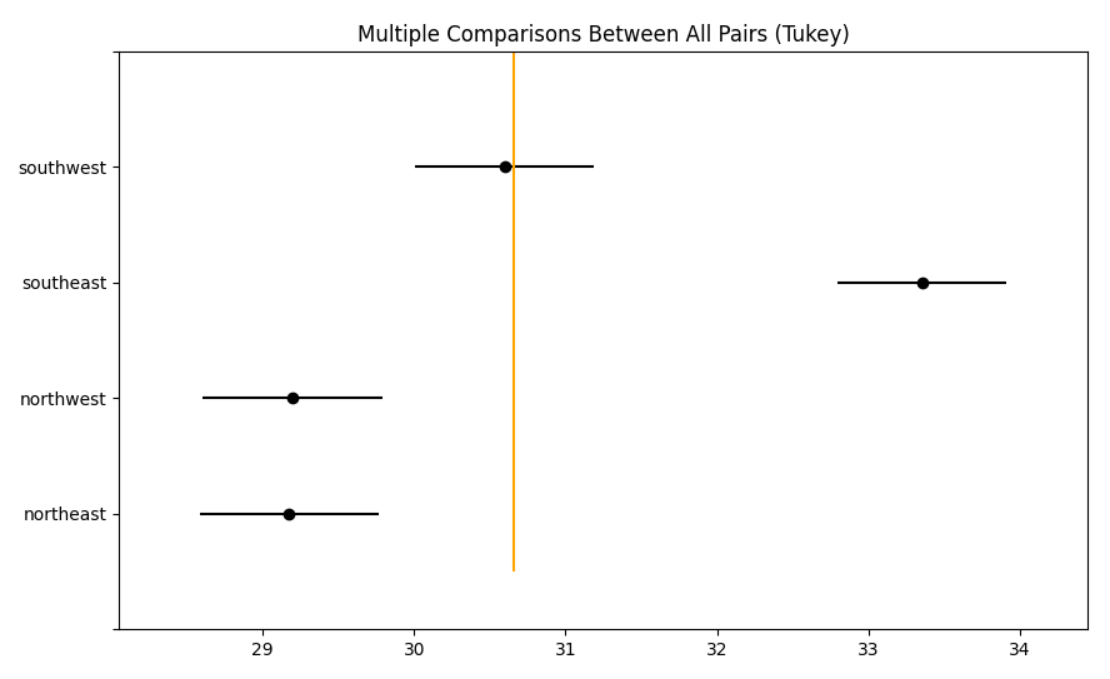


Рисунок 9 – Скриншот результата выполнения пятого задания

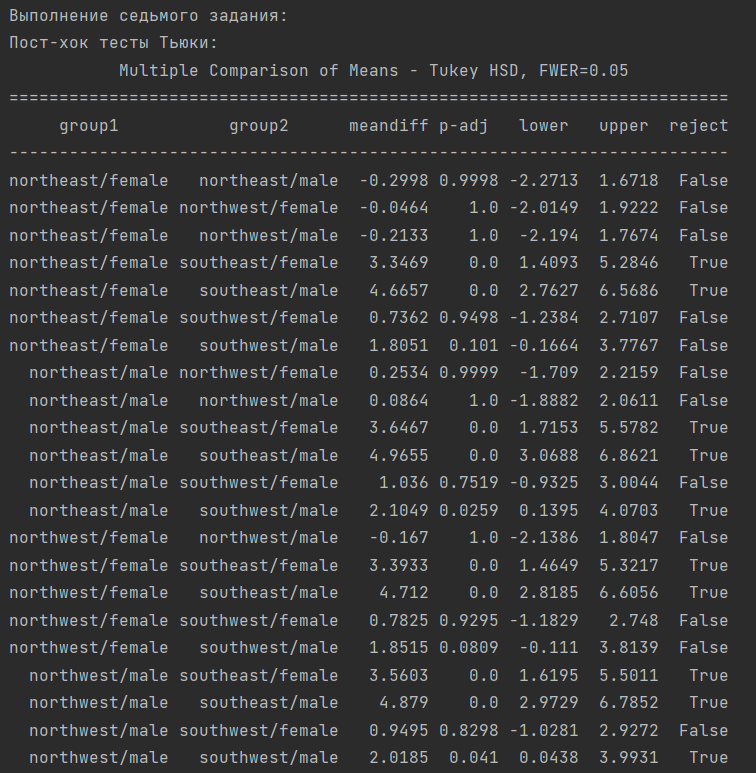


Рисунок 10 – Скриншот результата выполнения седьмого задания (Часть 1)

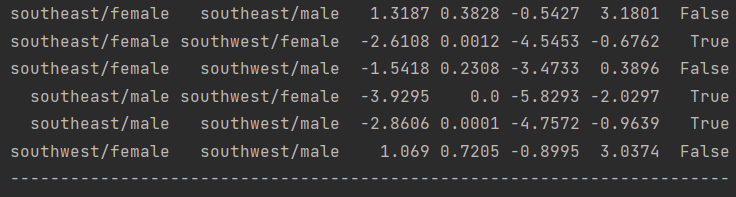


Рисунок 11 – Скриншот результата выполнения седьмого задания (Часть 2)

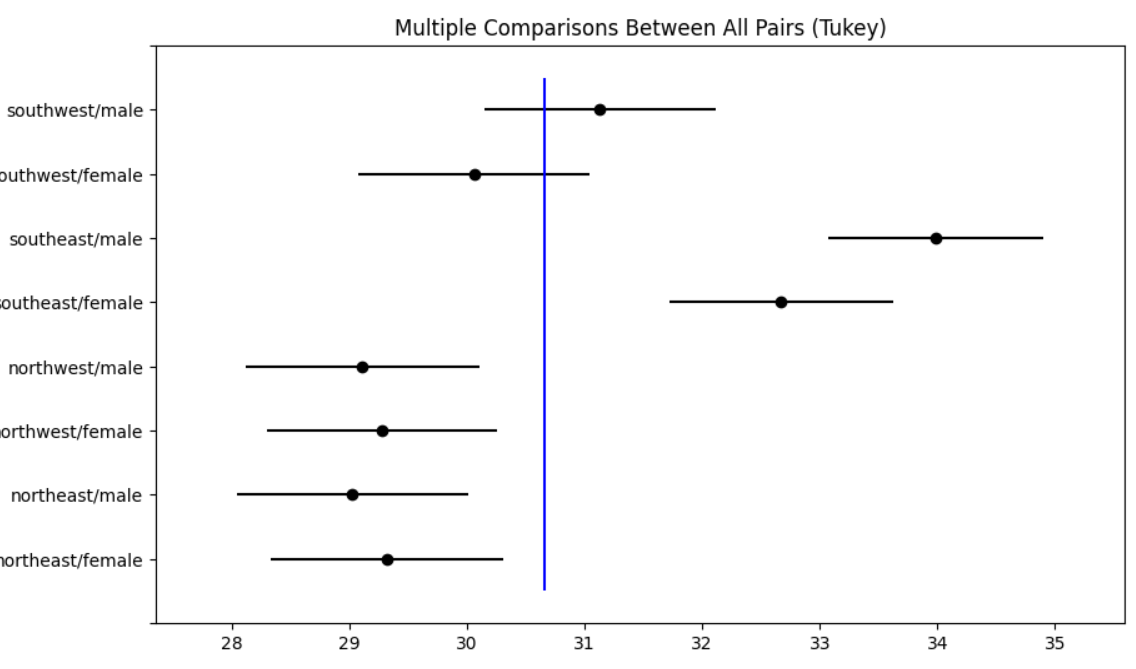


Рисунок 12 – Скриншот результата выполнения седьмого задания

**Вывод**

В результате выполнения данной практической работы были загружены данные из датасета, произведена предобработка датасета, выведен список уникальных регионов, выполнены однофакорный и двухфакторный ANOVA тесты несколькими способами, с помощью t критерия Стьюдента перебраны все пары, определена поправка Бонферонни, выполнены пост-хок тесты Тьюки, построены соответствующие графики, наконец, оформлен отчет о проделанной работе.