|  |
| --- |
|  |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА - Московский технологический университет»**  **РТУ МИРЭА** |

Институт Информационных Технологий

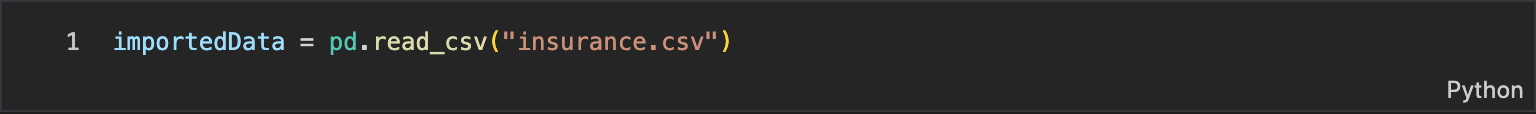
Кафедра Прикладной Математики (ПМ)

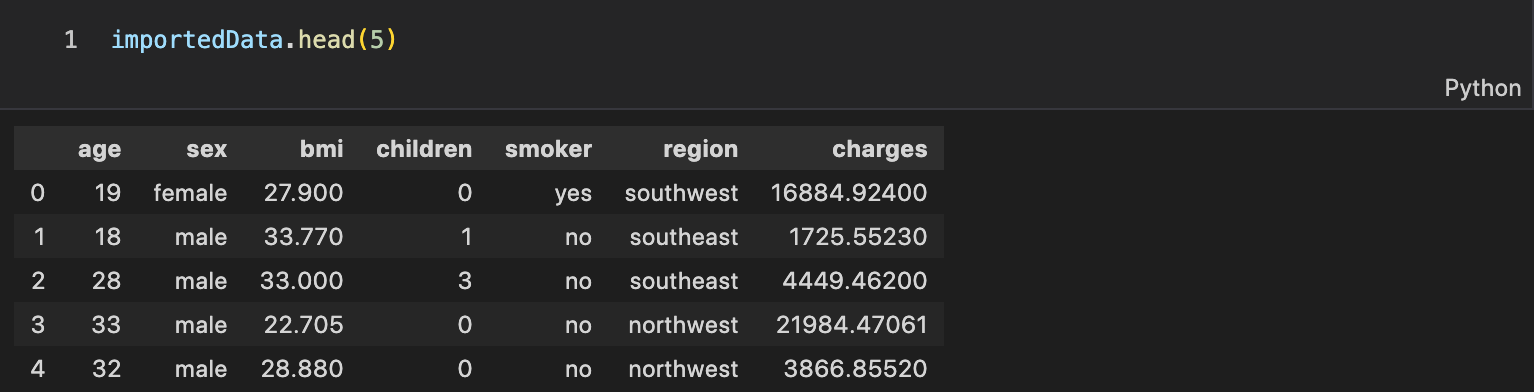
**ПРАКТИЧЕСКАЯ РАБОТА № 9**

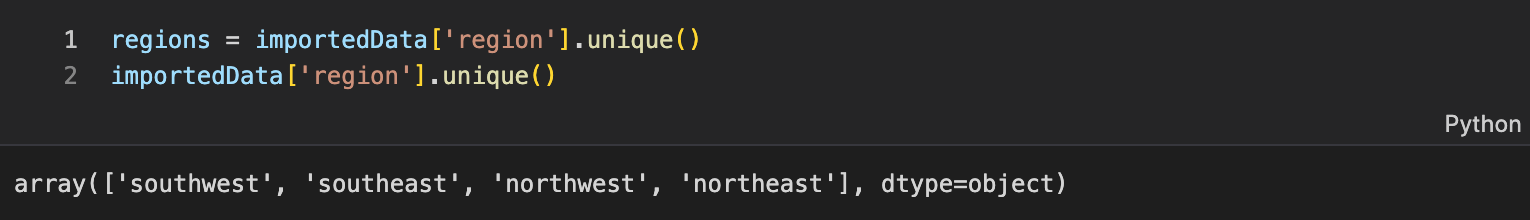
|  |  |
| --- | --- |
| Выполнил студент группы ИКБО-08-19  Борисов А.В. | (\_\_\_\_\_\_\_\_\_\_)  *подпись* |
| Принял Ассистент кафедры ПМ  Высоцкая А.А. | (\_\_\_\_\_\_\_\_\_\_)  *подпись* |
| Практическая работа выполнена | « » 2022 г. |
| «Зачтено» | « » 2022 г. |

Москва 2022

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



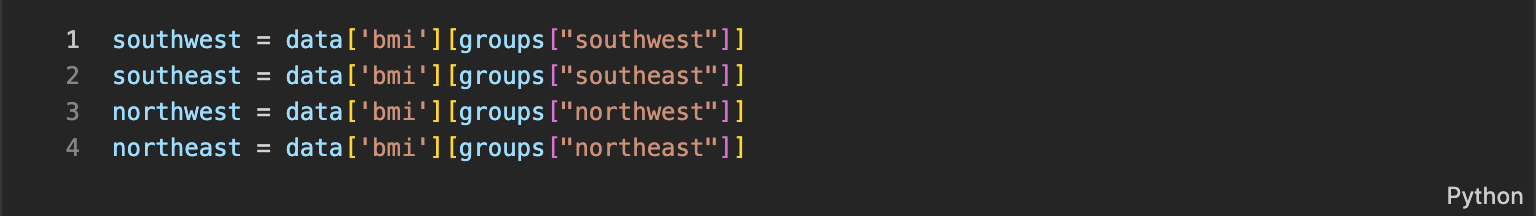


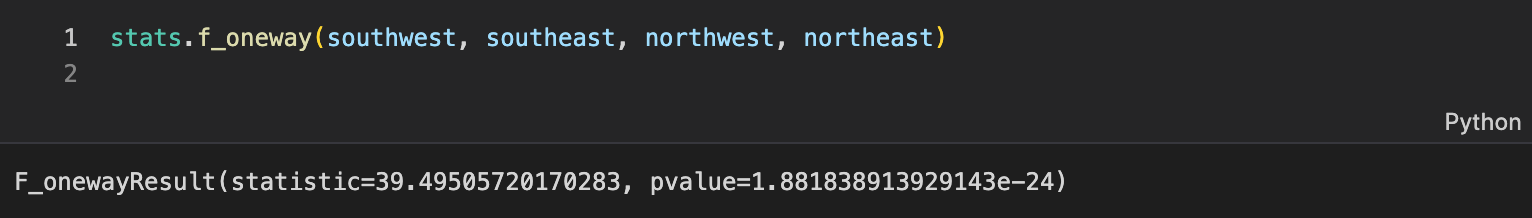


1. **Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя первый способ, через библиотеку Scipy.**



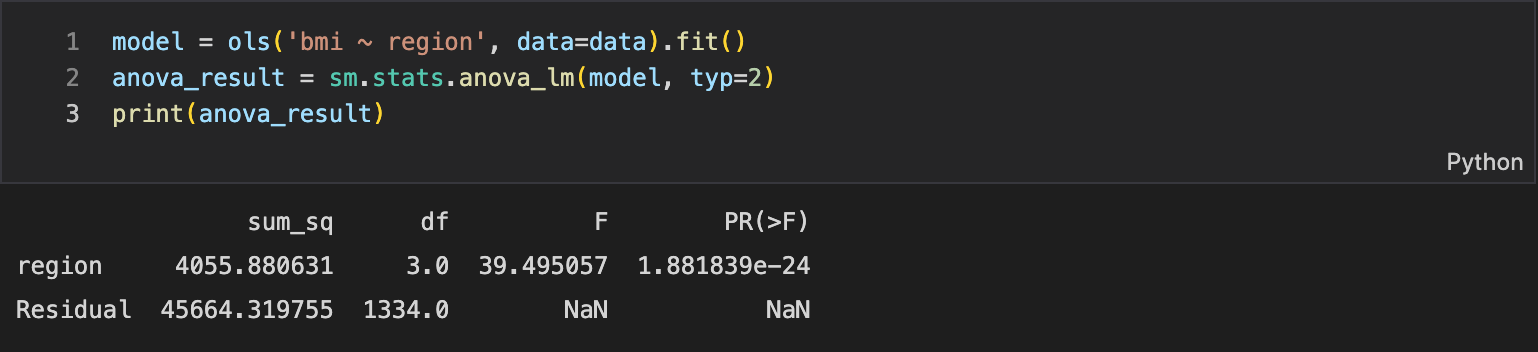






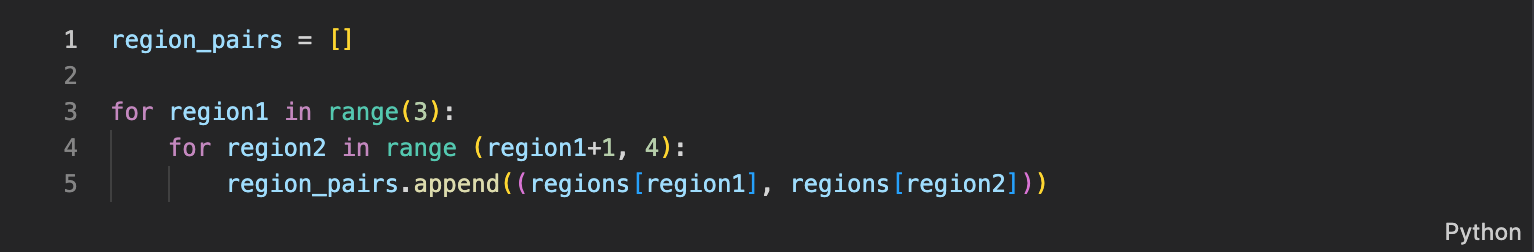
Поскольку p < 0.05 можно сделать вывод, что нулевую теорию (гласящую о том, что средние групп равны) можно отбросить, соответственно принимается альтернативная теория, которая говорит о различии средних. Отсюда следует вывод, что фактор 'region' имеет статистически значимое влияние на показатель 'bmi' данных групп.

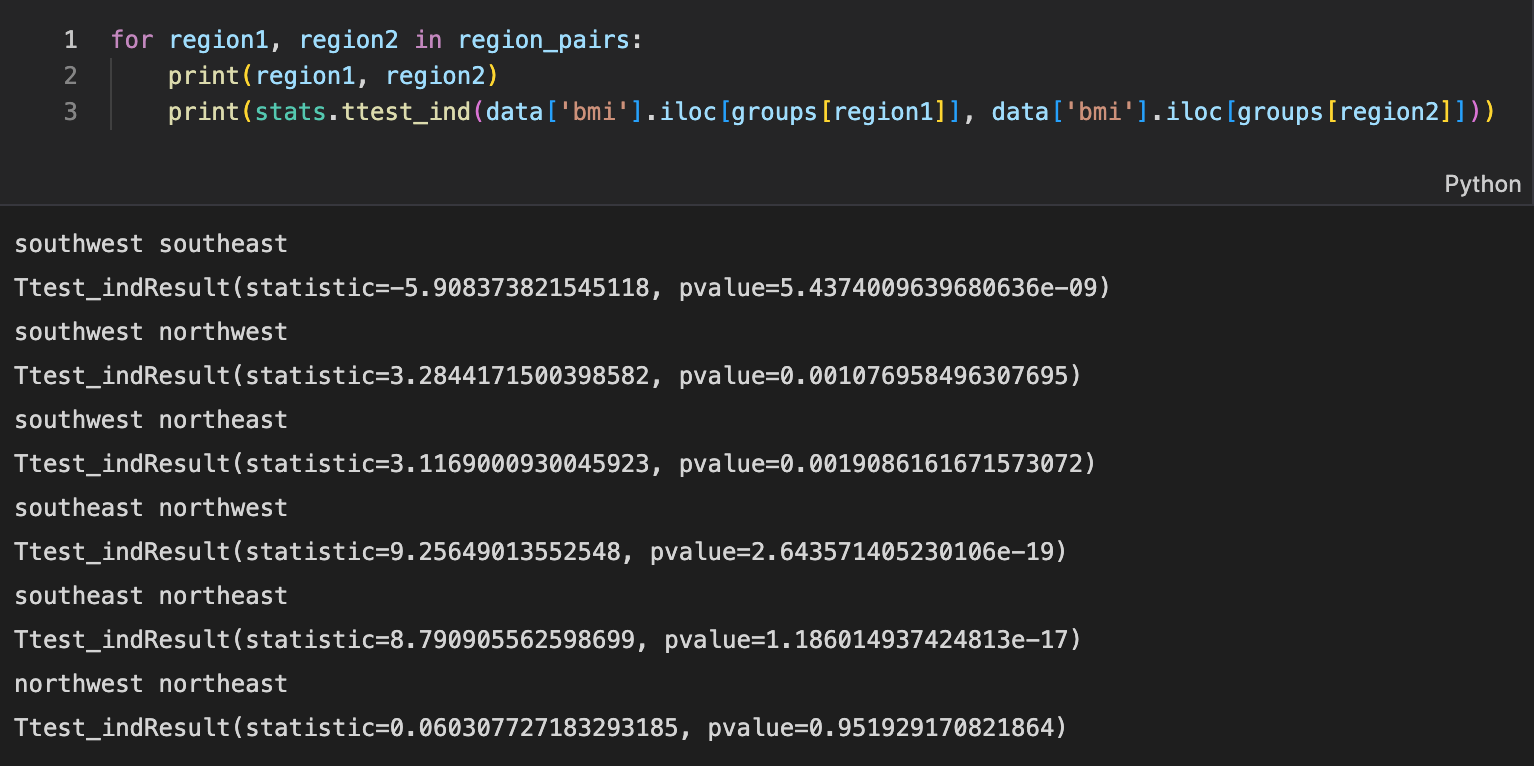
1. **Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя второй способ, с помощью функции anova\_lm() из библиотеки statsmodels.**



Вывод тот же самый, что и в предыдущем задании. р < 0.05...

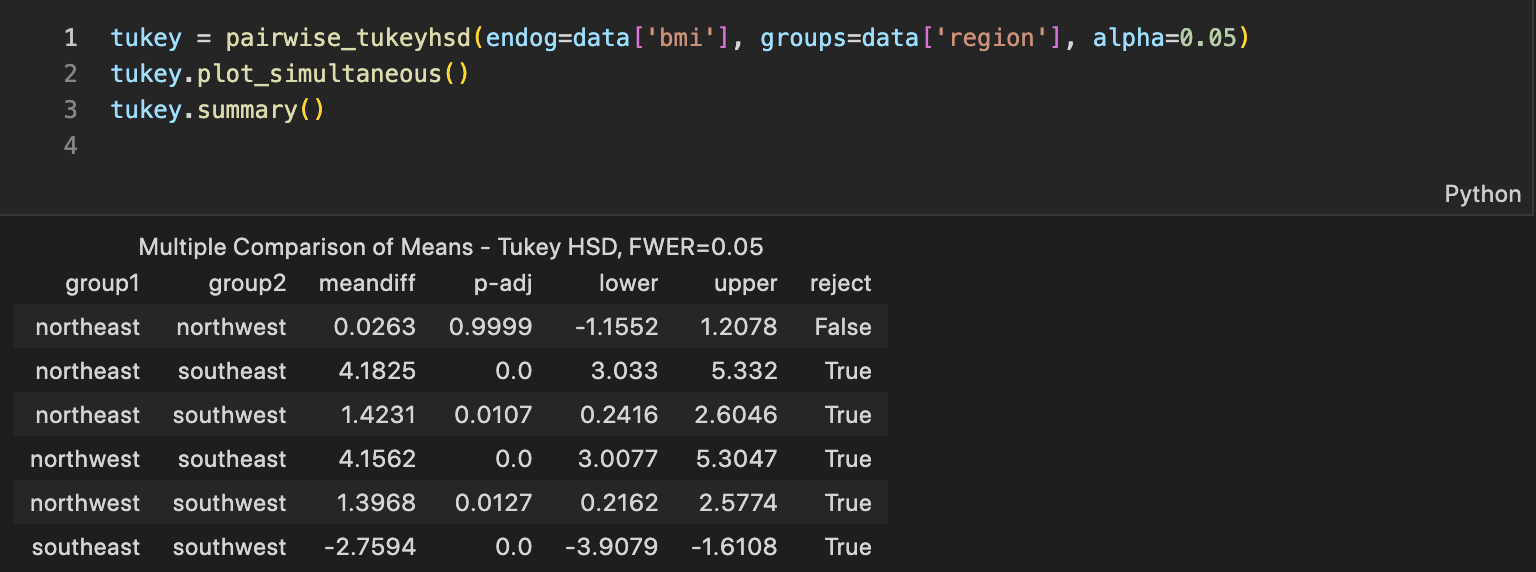
1. **С помощью t критерия Стьюдента перебрать все пары. Определить поправку Бонферрони. Сделать выводы.**

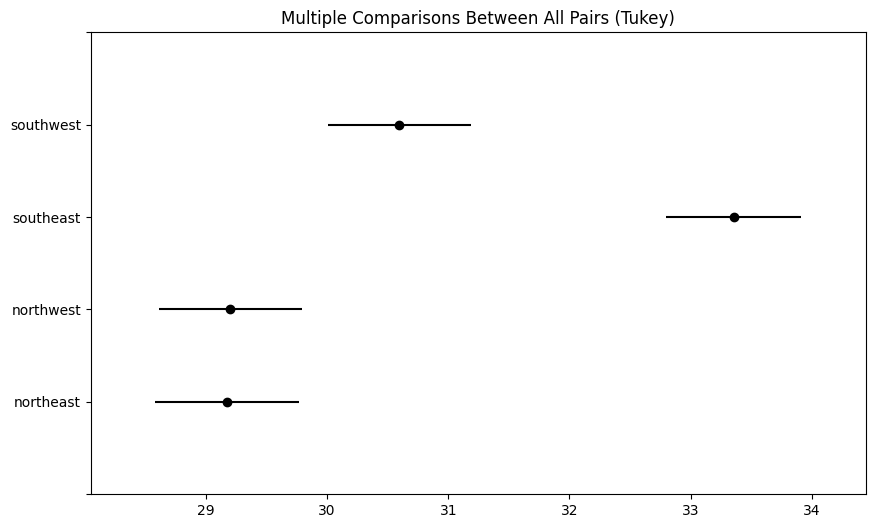




Из результатов данного теста можно сделать выводы, что все пары групп кроме northwest-northeast имеют статистически значимые различия.

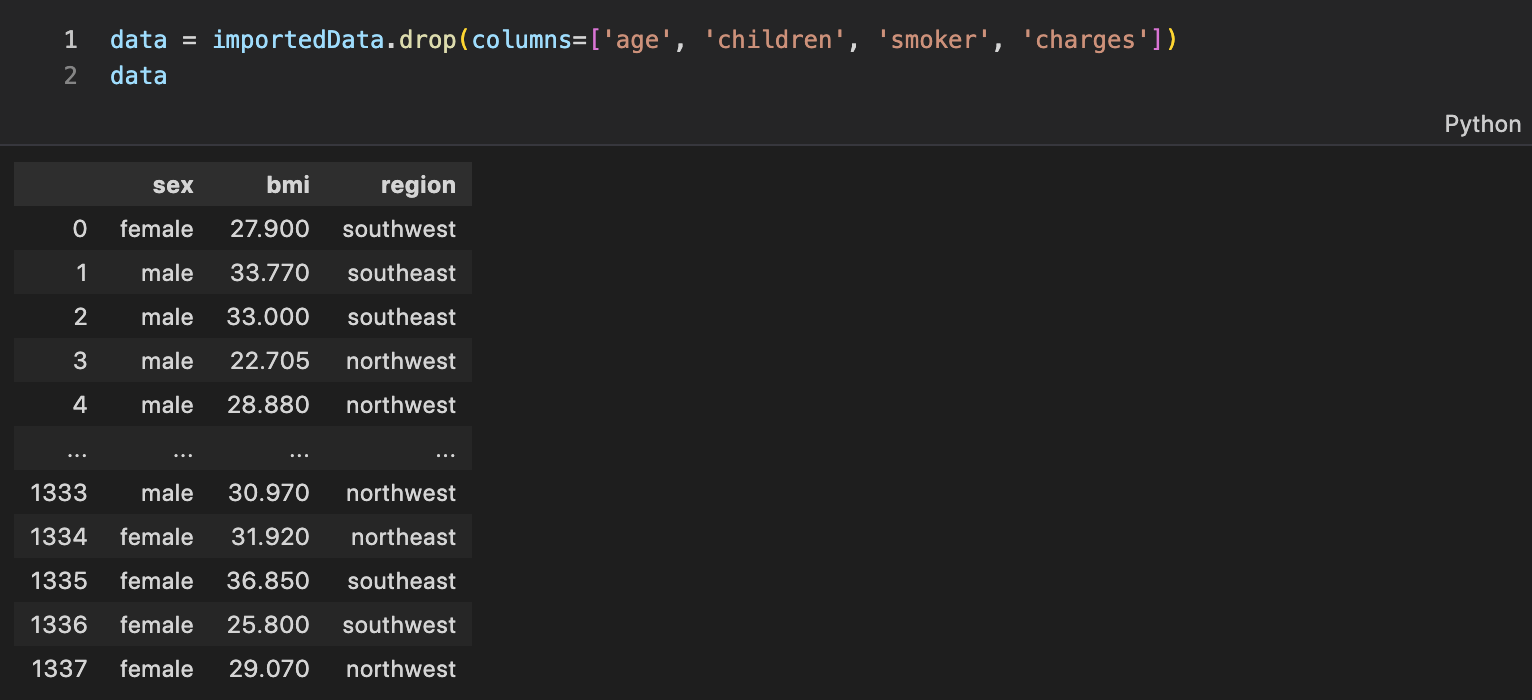
1. **Выполнить пост-хок тесты Тьюки и построить график.**

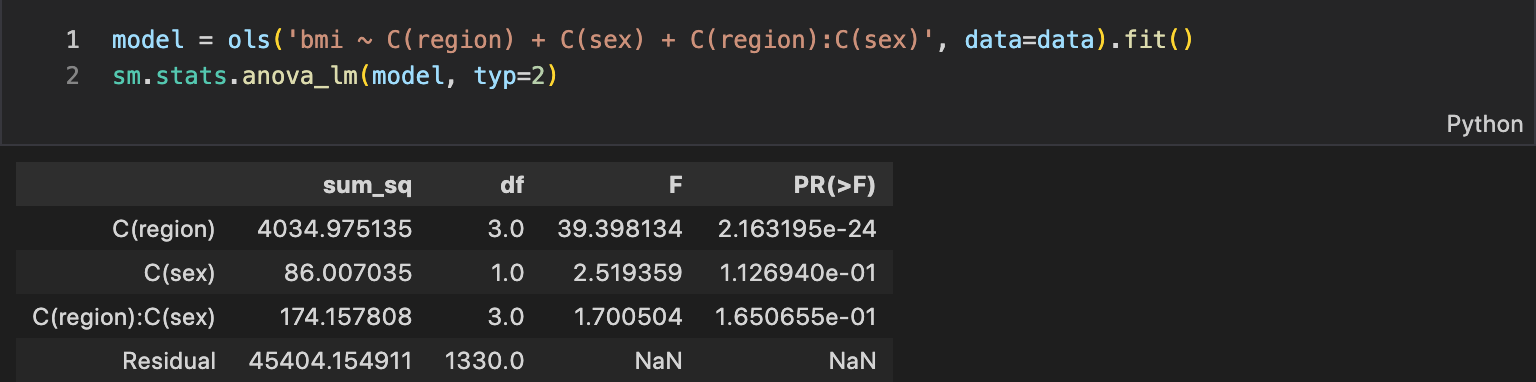




Как и в предыдущем тесте, все группы показывают различие в BMI, кроме пары northeast - northwest. График это подтверждает, доверительные интервалы этих групп перекрываются.

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

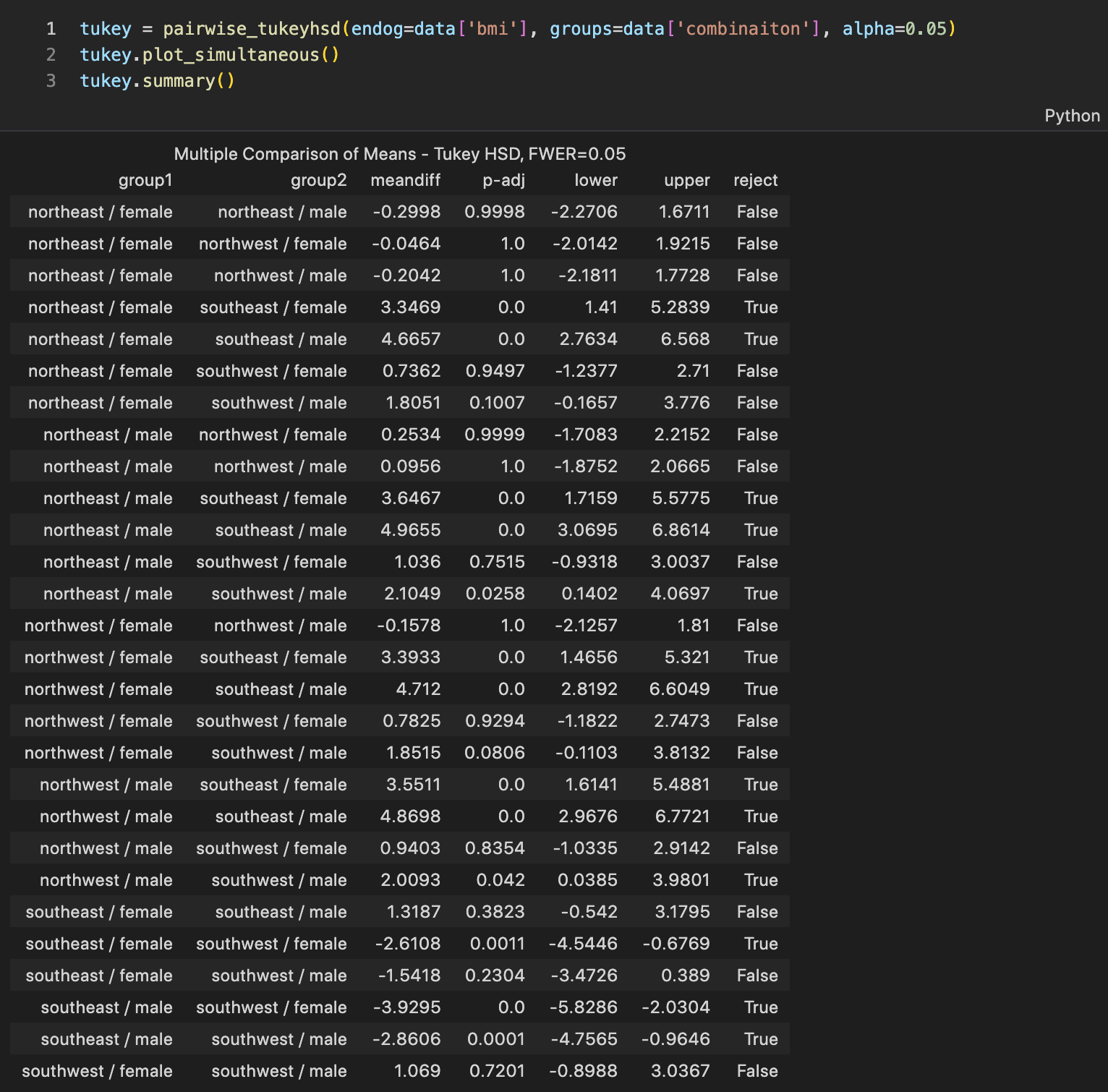




P-значения фактора region меньше 0.05, это значит что он оказывает статистически значимое влияние на значение BMI.

1. **Выполнить пост-хок тесты Тьюки и построить график.**





Из выполненного теста можно сделать выводы, что хоть и многие группы имеют статистически важные различия (о чем свидетельствует значение 'true' в колонке 'reject'), не все так однозначно, и к примеру у групп northwest и southwest есть пересекающиеся доверительные интервалы в многих подгруппах, что может говорить и о близости их значений. К слову группы из северных регионов очень похожи, и имеют больше общего, чем группы из южных, где southeast отличается от southwest больше, чем последний отличается от северных регионов.