

Домашняя работа №5

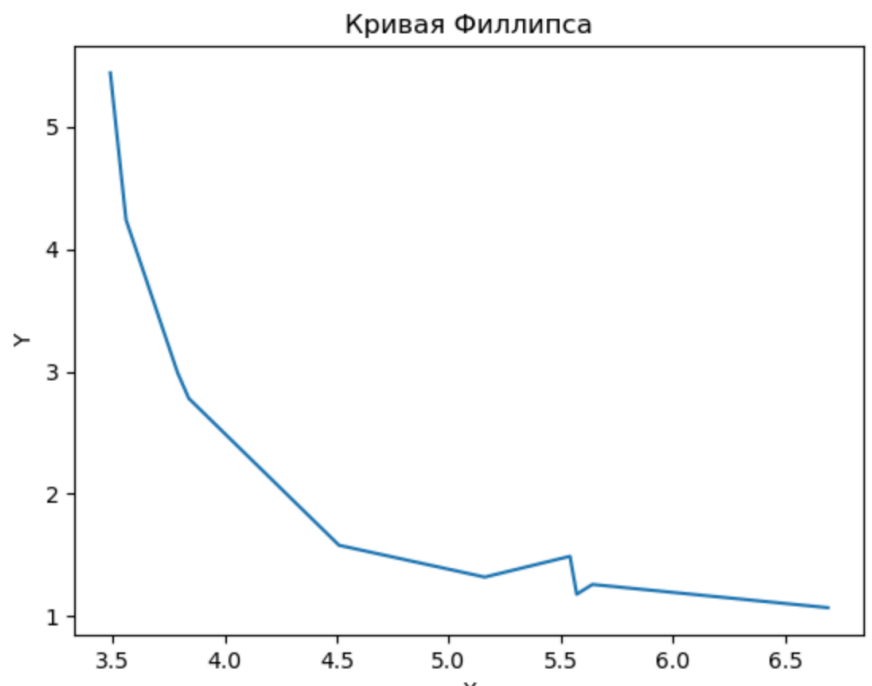
Проверка модели Филлипса на предмет наличия гомоскедастичности

Гомоскедастичность - это статистическое свойство данных, которое означает, что дисперсия ошибок модели не зависит от значений независимых переменных. Если данные гомоскедастичны, то они имеют постоянную дисперсию вдоль всего диапазона значений предикторов.

1. Тесты на гомоскедастичность

Для анализа второго условия Гаусса - Маркова я буду использовать приближённые данные инфляции и безработицы в США с 1960-1970.

	безработица	инфляция
0	5.54	1.49
1	6.69	1.07
2	5.57	1.18
3	5.64	1.26
4	5.16	1.32
5	4.51	1.58
6	3.79	2.99
7	3.84	2.78
8	3.56	4.24
9	3.49	5.44



1) Тест ранговой корреляции Спирмена.

Используем формулу r_s , предварительно проранжировав значения остатков и x .

```
p = 1 - 6*sum((abs(data1['Xr']-data1['Er'])*2))/(1000 - 10)
p
```

0.018181818181818188

Далее находим коэффициент ранговой корреляции и сравниваем его со значением в таблице значений t критерия Стьюдента

```
t = abs(p) * np.sqrt(8)/np.sqrt(1 - p**2)
t
0.051434449987363996
```

В данном случае коэффициент корреляции меньше 2,306 при уровне значимости равному 0,05 следовательно гетероскедастичность не доказана.

2)Тест Голдфелда-Куандта

Использование данного теста подразумевает упорядочивали остатков по возрастанию и последующее деление совокупности на 3 равные части. F статистика находится по формуле: сумма остатков в квадрате(первой части наблюдений) делённая на сумма остатков в квадрате(второй части наблюдений).

В питоне можно реализовать с помощью statsmodels:

```
sm.stats.diagnostic.het_goldfeldquandt(data1['Y'], sm.add_constant(data1['X']), drop=0.2)
(27.701896238658936, 0.1385588984371137, 'increasing')
```

Сравниваем F stat с табличкам значением(6,39) -> Fstat > F табл

Подтверждается наличие гетероскедастичности.

3) Тест Уайта

Тест подразумевает наличие нескольких зависимых переменных , тем не менее я решил его провести на заданной модели.

```
white_test = sm.stats.diagnostic.het_white(model.resid**2, model.model.exog )
labels = ['Test Statistic', 'Test Statistic p-value', 'F-Statistic', 'F-Test p-value']
print(dict(zip(labels, white_test)))
{'Test Statistic': 1.8415988425610652, 'Test Statistic p-value': 0.39820058379010265, 'F-Statistic': 0.7900562652630228, 'F-Test p-value': 0.49047592717498883}
```

F знач < F табл следовательно наличие гетероскедастичности не доказано.

4) Тест Бреуша-Пагана

Тест Бреуша-Пагана использует следующие нулевые и альтернативные гипотезы:

Нулевая гипотеза (H_0): присутствует гомоскедастичность.

Альтернативная гипотеза: (H_a): гомоскедастичность отсутствует

```
names = ['Lagrange multiplier statistic', 'p-value',  
'f-value', 'f p-value']  
test = sm.stats.diagnostic.het_breuschpagan(model.resid, model.model.exog)  
  
print(dict(zip(names, test)))
```

```
{'Lagrange multiplier statistic': 1.6418712690533743, 'p-value': 0.20006891581741104, 'f-value': 1.571520441387046,  
'f p-value': 0.24537998878398215}
```

Так как p value не меньше 0,05 не доказано присутствие гетероскеластичности, гипотеза не принимается.

2. Теоретическая часть(суть тестов)

1) Тест ранговой корреляции Спирмена используется, когда у нас есть две ранжированные переменные, и мы хотим увидеть имеет ли при увеличении одной переменной тенденцию другая переменная увеличиваться или уменьшаться. Мы также используем ранговую корреляцию Спирмена, если у нас есть одна переменная измерения и одна ранжированная переменная; в этом случае мы конвертируем переменную измерения в ранги и используем ранговую корреляцию Спирмена для двух наборов рангов.

2) Тест Голдфелда-Куандта сравнивает дисперсии двух подгрупп; один набор высоких значений и один набор низких значений. Если отклонения различаются, тест отвергает нулевую гипотезу о том, что дисперсии ошибок не являются постоянными. Статистика для этого теста представляет собой соотношение среднеквадратичных остаточных ошибок для регрессий на двух подмножествах данных. Это соответствует F-тесту на равенство дисперсий.

3) Тест Уайта зачастую используется если в нашем наборе данных много объясняющих переменных. Тест Уайта — это асимптотический тест, который предназначен для использования на больших выборках. Одна из проблем с тестом Уайта заключается в том, что он может дать значимый результат, даже если дисперсии ошибок равны. Это происходит потому, что модель является общей и может обнаружить другие проблемы в наших данных.

4) Тест Бреуша-Пагана заключается в:

1. Рассчёте квадратов остатков модели.
2. Подборе новой модели регрессии, используя квадраты остатков в качестве значений ответа.
3. В конечном итоге рассчитываем статистику критерия как $n \cdot R^2$, где: n : Общее количество наблюдений, а R^2 : значения новой модели регрессии, в которой в качестве значений ответа использовались квадраты остатков.

Источники информации

- 1) <https://www.statisticshowto.com/white-test/>
- 2) <https://spureconomics.com/breusch-pagan-test-for-heteroscedasticity/>
- 3) <http://www.biostathandbook.com/spearman.html#:~:text=When%20to%20use%20it,tends%20to%20increase%20or%20decrease.>