

## Задание 2

Юневич Дарья, 312 группа.

### Описание задачи

1. Считать данные из `Данные.xlsx`. Проверить является ли ряд стационарным в широком смысле. Это можно сделать двумя способами:
  - i. Провести визуальную оценку, отрисовав ряд и скользящую статистику (среднее, стандартное отклонение). Построить график, на котором будет отображен сам ряд и различные скользящие статистики.
  - ii. Провести тест Дики - Фуллера.

Сделать выводы из полученных результатов. Оценить достоверность статистики.

2. Разложить временной ряд на тренд, сезонность, остаток в соответствии с аддитивной, мультипликативной моделями. Визуализировать их, оценить стационарность получившихся рядов, сделать выводы.
3. Проверить является ли временной ряд интегрированным порядка  $k$ . Если является, применить к нему модель ARIMA, подобрав необходимые параметры с помощью функции автокорреляции и функции частичной автокорреляции. Выбор параметров обосновать. Отобрать несколько моделей. Предсказать значения для тестовой выборки. Визуализировать их, посчитать  $r2\_score$  для каждой из моделей. Произвести отбор наилучшей модели с помощью информационного критерия Акаике. Провести анализ получившихся результатов.

### Описание алгоритма

*Определения:*

- Временной ряд – это последовательность значений, описывающих протекающий во времени процесс, измеренных в последовательные моменты времени, обычно через равные промежутки.
- Сезонность – компонента, которая описывает циклические изменения временного ряда.
- Тренд - компонента, описывающая долгосрочное изменение уровня ряда.
- Остатком временного ряда называется разница между предсказанным и наблюдаемым значением.
- Автокорреляция — статистическая взаимосвязь между последовательностями величин одного ряда. Это один из самых важных коэффициентов в анализе временного ряда. Чтобы посчитать автокорреляцию, используется корреляция между временным рядом и её сдвинутой копией от величины временного сдвига. Сдвиг ряда называется лагом.
- Стационарный ряд - ряд, в котором свойства не зависят от времени. При таких компонентах ряда, как тренд или сезонность, ряд не является стационарным. Есть два вида стационарности - строгая и слабая (стационарность в широком смысле). Слабая стационарность заключается в постоянной дисперсии и в постоянности среднего значения ряда.

В рамках визуальной оценки обращают внимание на наличие тренда у графика. Если тренд есть - ряд не стационарный, тренда нет - ряд стационарен.

Для более точной оценки стационарности используют тест Дики-Фуллера, который опирается на статистический вывод. В данном тесте за нулевую гипотезу берется утверждение, что

процесс не стационарен, а за альтернативную, что стационарен. Пороговое значение принято брать равное 0,05 (5%). В результате наш вывод будет зависеть от следующих суждений: если вероятность (p-value) для данных существенно выше 0,05, то мы не отвергаем нулевую гипотезу и процесс считаем нестационарным, иначе, если p-value меньше 0,05 – считаем ряд стационарным.

\*Для вычисления p-value используем функцию `adfuller()` из библиотеки `statsmodels`.

Аддитивная модель разложения временного ряда - *тренд* + *сезонность* + *остаток*.

Для получения функций тренда, сезонности и остатка аддитивной модели, воспользуемся функцией `seasonal_decompose` с параметром `model = 'additive'`.

Мультипликативная модель разложения временного ряда - *тренд* \* *сезонность* \* *остаток*.

Для получения функций тренда, сезонности и остатка мультипликативной модели, воспользуемся функцией `seasonal_decompose` с параметром `model = 'multiplicative'`.

- Интегрированный временной ряд — нестационарный временной ряд, разности некоторого порядка от которого являются стационарным временным рядом
- Интегрированный ряд порядка k – ряд, в котором разности k-го порядка  $\Delta^k x(t)$  являются стационарными, а разности меньшего порядка (и сам временной ряд) не являются стационарными рядами.
- ARIMA (autoregressive integrated moving average model) - интегрированная модель авторегрессии скользящего среднего - модель анализа временных рядов. Это расширение моделей ARMA для нестационарных временных рядов.
- ARMA- математическая модель, используемая для анализа и прогнозирования стационарных временных рядов.

Для построения модели ARIMA нужно найти коэффициенты p и q, это можно сделать с помощью функция автокорреляции и функция частичной автокорреляции.

- Коэффициент детерминации — это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью зависимости, то есть объясняющими переменными. (наилучшим значением `r2_score` считается 1)
- Информационный критерий — применяемая в эконометрике мера относительного качества эконометрических (статистических) моделей, учитывающая степень «подгонки» модели под данные с корректировкой (штрафом) на используемое количество оцениваемых параметров. Критерии основаны на компромиссе между точностью и сложностью модели. Критерии различаются тем, как они обеспечивают этот баланс.
- Информационный критерий Акаике (AIC) — критерий, применяющийся исключительно для выбора из нескольких статистических моделей.

## Описание программы

В программе были последовательно получены результаты, требуемые в описании задания. В коде в основном использовались библиотечные функции такие как: ARIMA – для модели ARIMA, `plot_acf` и `plot_pacf` – для нахождения корреляций, `seasonal_decompose` – для разложения на тренд, сезонность и остаток, а также другие функции. Для визуализации данных была использована библиотека `seaborn`.

## Инструкция к запуску

Необходимое ПО: Windows 7

Запустить программу можно в google colab, пройдя по ссылке, загрузив предварительно файлы – Answers.xlsx и Data.xlsx

<https://colab.research.google.com/drive/13BZ5BOiNiUxmN-y2HHVHzhO6m34Mard9?usp=sharing>