

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. Ломоносова
Факультет Вычислительной Математики и Кибернетики

ЗАДАНИЕ №2

ОТЧЕТ

О выполненном задании

группы студентов 311 и 312 учебной группы факультета ВМК МГУ

Александр Борисов

Иван Наумов

Охим Елена

Максим Гончаренко

Москва, 2020 г.

Содержание

1	Постановка задачи и подход к ее решению	2
1.1	Цели и задачи практической работы	2
2	Теоритические основы	3
2.1	Анализ временного ряда	3
2.2	Этапы решения задачи:	4
3	Необходимое ПО	5
4	Вклад каждого участника	6

1 Постановка задачи и подход к ее решению

Постановка задачи состоит в проведении анализа некоторого временного ряда и попытке предсказать значения для последующих месяцев. В рамках данного задания мы: научимся правильно считывать данные и их визуализировать; поймем как определять свойства временных рядов и познакомимся с различными моделями для предсказания значений; оценим качество полученных моделей..

1.1 Цели и задачи практической работы

- Знакомство с правильным считыванием данных и их визуализацией.
- Определение свойств временных рядов и знакомство с различными моделями для предсказания значений.
- Оценивание качества полученных моделей.
- Продолжение изучения языка Python, знакомство с такими библиотеками, как Pandas(работа с данными), statsmodels(работа с различными статистическими моделями, в том числе и временными рядами).

2 Теоритические основы

2.1 Анализ временного ряда

Временной ряд — это последовательность значений признака y , измеряемого через постоянные временные интервалы.

Анализ временного ряда — это совокупность математико-статистических методов анализа, предназначенных для выявления структуры временных рядов и для их прогнозирования.

Тренд — это плавное долгосрочное изменение уровня ряда. Эту характеристику можно получить, наблюдая ряд в течение достаточно долгого времени.

Сезонность — это циклические изменения уровня ряда с постоянным периодом.

Цикл — это изменение уровня ряда с переменным периодом.

Ошибка — это непрогнозируемая случайная компонента ряда.

Количественной характеристикой сходства между значениями ряда в соседних точках является автокорреляционная функция (или просто автокорреляция), которая задаётся следующим соотношением:

$$E((y_t - E y)(y_{t+\tau} - E y)) / D$$

Анализировать величину автокорреляции при разных значениях лагов удобно с помощью графика, который называется коррелограммой. По оси ординат на нём откладывается автокорреляция, а по оси абсцисс — размер лага

Временной ряд называется **стационарным**, если $\forall s$ - ширина окна распределение y_t, \dots, y_{t+s} не зависит от t , т.е. его свойства не зависят от времени.

Скользящая статистика — это общее название для семейства функций, значения которых в каждой точке определения равны среднему значению исходной функции за предыдущий период. Скользящая статистика обычно используется с данными временных рядов для сглаживания краткосрочных колебаний и выделения основных тенденций или циклов.

Аддитивная модель имеет вид: $Y = T + S + E$ где T - компонента тренда, S - компонента сезонности, E - случайная компонента.

Мультипликативная модель имеет вид: $Y = TSE$ где T - компонента тренда, S - компонента сезонности, E - случайная компонента

Временной ряд называется **интегрированным порядка k** , если его разности порядка k образуют стационарный ряд.

2.2 Этапы решения задачи:

Этап 1

Для начала требуется проверить ряд на стационарность. Один из способов - это визуальная оценка путем рисования ряда и скользящей статистики.

Этап 2

Разложение временного ряда на тренд и сезонность.

Этап 3

Так как ряд является интегрированным порядка 1, можно применить к нему модель ARIMA. Выбор параметров. Предсказание значений для тестовой выборки. Визуализация. Отбор наилучшей модели с помощью информационного критерия Акаике.

3 Необходимое ПО

Библиотеки:

- datetime
- matplotlib
- seaborn
- pandas
- numpy
- statsmodels
- warnings
- pylab
- sklearn
- itertools

4 Вклад каждого участника

- Александр Борисов сделал разложение временного ряда на тренд, сезональность и шум.
- Иван Наумов полностью собрал программу, реализовал тест Дики - Фуллера, работал с github.
- Охим Елена построила прогнозирующей модели и оформила readme.
- Максим Гончаренко сделал проверку ряда на стационарность посредством визуализации статистик.