НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Факультет прикладної математики

Кафедра прикладної математики

Курсовий проект

із дисципліни «Алгоритми і системи комп’ютерної математики»

На тему

«Прогнозування кількості хворих на COVID-19»

Етап №2

Виконав: Керівник:

студент групи КМ-93 доцент

Костенко О. А. Олефір О. С.

Київ — 2022

Опис обраного математичного методу

Підхід авторегресійного інтегрованого ковзного середнього (англ. autoregressive integrated moving average, ARIMA) до часових рядах полягає в тому, що в першу чергу оцінюється стаціонарность ряду. Різними тестами виявляються наявність поодиноких коренів і порядок інтегрованості часового ряду (зазвичай обмежуються першим або другим порядком). Далі, при необхідності, (якщо порядок інтегрованості більше нуля) ряд перетворюється взяттям різниці відповідного порядку і вже для перетвореної моделі будується деяка ARMA-модель, оскільки передбачається, що отриманий процес є стаціонарним, на відміну від вихідного нестаціонарного процесу (разностно-стаціонарного або інтегрованого процесу порядку d).

ARIMA - модель і методологія аналізу часових рядів. Є розширенням моделей ARMA для нестаціонарних часових рядів, які можна зробити стаціонарними взяттям різниць деякого порядку від вихідного часового ряду (так звані інтегровані або різносно-стаціонарні часові ряди). Модель 𝐴𝑅𝐼𝑀𝐴(𝑝,𝑑,𝑞) означає, що різниці часового ряду порядку d належать моделі 𝐴𝑅𝑀𝐴(𝑝,𝑞).

Модель 𝐴𝑅𝐼𝑀𝐴(𝑝,𝑑,𝑞) для нестаціонарного часового ряда 𝑋𝑡 має вигляд:,

де – це зпрогнозоване значення на період 𝑡;

– константа, зазвичай для спрощення рівняється нулю;

– параметри моделі, коефіцієнти авторегресії;

– параметри моделі, ковзного середнього;

– оператор різниці часового ряду порядку 𝑑 (послідовне взяття 𝑑 раз різниць першого порядку - спочатку від тимчасового ряду, потім від отриманих різниць першого порядку, потім від другого порядку і т.д.)

– білий шум.

Також, дана модель інтерпретується як 𝐴𝑅𝑀𝐴(𝑝+𝑑,𝑞) модель з 𝑑 одиничними розв'язками. При 𝑑 = 0 маємо звичайну ARMA модель.

ARIMA-моделі дозволяють моделювати інтегровані або разностностаціонарні часові ряди (DS-ряди, diference stationary). Часовий ряд називається інтегрованим порядку 𝑘, якщо різниці ряду порядку ∆𝑘𝑥𝑡, тобто є стаціонарними, в той час як різниці меншого порядку (включаючи нульового порядку, тобто сам тимчасової ряд) не є стаціонарними щодо деякого тренда рядами (TS-рядами, trend stationary ). В зокрема 𝐼(0) – це стаціонарний процес. Порядок інтегрованості часового ряду і є порядок 𝑑 моделі.

**Методологія Бокса-Дженкінса.** Методологія Бокса-Дженкінса підбору ARIMA моделі для даного ряду спостережень складається з 5 кроків.

**Крок 1.** Отримання стаціонарного ряду. Ми тестуємо ряд на стаціонарність, використовуючи описані вище методи: візуальний аналіз графіка, візуальний аналіз ACF і PACF, тести на одиничні коріння. Якщо виходить стаціонарний ряд, то переходимо до наступного пункту, якщо немає стаціонарності ряду, то застосовуємо оператор взяття послідовної різниці і повторюємо тестування. Напрактиці послідовна різниця береться, як правило, не більше двох разів.

**Крок 2.** Після того, як отримано стаціонарний часовий ряд, будуються його вибіркові ACF і PACF, які, як було показано вище, є своєрідними «відбитками пальців» ARMA(p, q) процесу і дозволяють сформулювати кілька гіпотез про можливі порядки авторегресії 𝑝 і змінного середнього 𝑞. Зазвичай рекомендується використовувати моделі можливо більш низького порядку, як правило, з 𝑝 𝑞 3 (якщо немає сезонної компоненти).

**Крок 3.** Для кожної з обраних на першому етапі моделей оцінюються їх параметри і обчислюються залишки.

**Крок 4.** Кожна з моделей перевіряється, наскільки вона відповідає даним. З моделей, адекватних даними, вибирається найпростіша модель, тобто з найменшою кількістю параметрів.

**Крок 5.** Прогнозування. Після того, як обрана модель, можна будувати прогноз на один або кілька кроків за часом і оцінювати довірчі інтвервали прогнозованих значень.

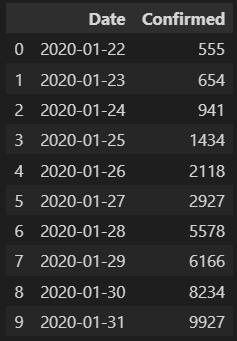
Модель S𝐴𝑅𝐼𝑀𝐴(𝑝,𝑑,𝑞, P, D, Q) - seasonal autoregressive integrated moving average, дуже схожа на модель *𝐴𝑅𝐼𝑀𝐴*(𝑝,𝑑,𝑞), за винятком того, що вона бере до уваги сезонність даних, з чого випливає додатковий набор компонентів авторегресії та ковзного середнього, що компенсуються частотою сезонності:

Модель S𝐴𝑅𝐼𝑀𝐴X(𝑝,𝑑,𝑞, r, P, D, Q), окрім сезонності, також враховує й екзогенні змінні, іншими словами, використовує зовнішні данні в прогнозі:

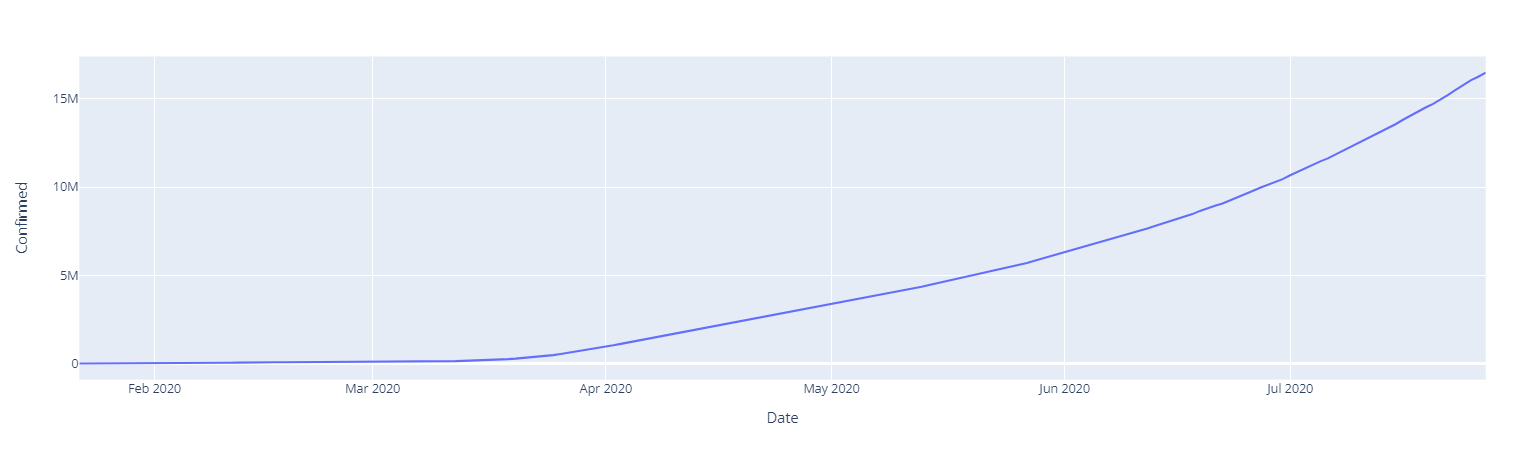
Контрольний приклад

Завантажимо дані про COVID-19 з файлу .csv. Залишивши в ньому тільки інформацію про дату та кількість хворих, отримуємо часовий ряд кількості хворих на COVID-19.

Розглянемо вхідні дані. Перші 10 записів файлу:



Побудова завантажених даних:



Для прогнозування кількості хворих на COVID-19 використаємо бібліотеку *statsmodels*, а саме модель *SARIMAX.* Розділимо дані на тренувальну та тестову вибірки з відношенням 70:30 та навчимо модель на тренувальних даних.

Результат прогноз моделі з урахуванням тестової виборки:

