МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

Кафедра прикладної математики та моделювання складних систем

Допущено до захисту
Завідувач кафедри ПМ та МСС
_____ к.ф.м.н., доцент
Коплик І. В.

«___» _____ 20__р.

КОМПЛЕКСНА КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня «бакалавр»

спеціальність 113 «Прикладна математика»

освітньо-професійна програма «Прикладна математика»

тема роботи «**МОДЕЛЮВАННЯ ПОШИРЕННЯ ЕПІДЕМІЇ COVID-19 В КРАЇНАХ ЄВРОПИ**»

Виконавець

Студент факультету ЕлІТ

Татаренко М. Д. _____

Науковий керівник

к.е.н., доцент

Маринич Т.О. _____

Суми – 2020

СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

Факультет **електроніки та інформаційних технологій**
Кафедра **прикладної математики та моделювання складних систем**

Рівень вищої освіти **перший**

Галузь знань **11 Математика та статистика**
Спеціальність **113 Прикладна математика**

Освітня програма **освітньо-професійна «Прикладна математика»**

ЗАТВЕРДЖУЮ
Завідувач кафедрою ПМтаМСС
к.ф.м.н., доцент Коплик І. В. _____
«___» _____2020 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ ВИЩОЇ ОСВІТИ ТАТАРЕНКО МИХАЙЛО ДЕНИСОВИЧ

Тема роботи: «МОДЕЛЮВАННЯ ПОШИРЕННЯ ЕПІДЕМІЇ COVID-19 В КРАЇНАХ ЄВРОПИ»

Керівник роботи: Маринич Тетяна Олександрівна, к.е.н., доцент

затверджено наказом по факультету ЕлІТ від «__» ___2020 р. № _____
Термін подання роботи студентом «___»_____2020 р.

Вихідні данні до роботи :

Зміст розрахунково-пояснювальної записки (перелік питань, що їх належить розробити):

Перелік графічного матеріалу: 15 рисунків

Дата видачі завдання «__» _____2020

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Термін виконання роботи	Примітка
1	Збір інформації щодо теми кваліфікаційної роботи		Звіт
2	Оформлення літературного огляду в звіті		Звіт
3	Розробка алгоритму і написання моделі		Мат. модель
4	Розробка програми мовою програмування R		Програма
5	Оформлення звітної документації з кваліфікаційної роботи		Звіт

Здобувач вищої освіти _____ Татаренко М. Д.

Керівник роботи _____ Маринич Т. О.

РЕФЕРАТ

Назва документа: звіт з кваліфікаційної роботи на тему «Моделювання поширення епідемії COVID-19 в країнах Європи».

Ключові слова: COVID-19, АНАЛІЗ ДАНИХ, СТАТИСТИКА, SIR МОДЕЛЬ, ВІЗУАЛІЗАЦІЯ, МОДЕЛЮВАННЯ, ПРОГНОЗУВАННЯ.

Короткий зміст документа та основні висновки: цей документ є кваліфікаційною роботою на здобуття освітнього ступеня «бакалавр» за спеціальністю 113 «Прикладна математика» на тему «Моделювання поширення епідемії COVID-19 в країнах Європи». Завданням кваліфікаційної роботи є підбір типу моделі та оптимальних значень параметрів, прогнозування показників поширення вірусу COVID-19 в Україні та країнах Європи. Об’єктом дослідження роботі є поширення епідемії COVID-19 та порівняння тенденцій поширення вірусу в Україні та Європі. Робота містить аналіз галузі дослідження, огляд наукових робіт на тему поширення вірусів, вивчення статистичних та математичних методів моделювання, висновки щодо проведеної роботи.

Кількість сторінок: 28.

Кількість рисунків: 15.

Кількість використаних джерел: 26.

Кількість додатків: 1.

ЗМІСТ

ВСТУП

Пандемія захворюваності вірусом COVID-19 охопила понад 200 країн та континентів та змінила наше уявлення про захищеність та перспективи людства у майбутньому. Більш ніж 5.5 млн підтверджених випадків зараження коронавірусом. Коли чуєш слово «пандемія», то уявляєш масштабні захворювання, такі як холера, іспанка, чума або грип. І з недавнього часу до такого переліку приєднався і COVID-19. Отже, на цей момент у світі налічується три пандемії: холера, ВІЛ-інфекції та COVID-19. Виходячи з останніх даних про темпи розповсюдження, кількість захворювань та смертей, COVID-19 вже на даний можна назвати «пандемією двадцять першого століття».

З появою комп'ютера, прогнозування та моделювання розповсюдження хвороби стало невід'ємною частиною певних наукових робіт. Це допомагає оцінити масштабність, належність вводити або скасовувати карантинні заходи, порівняти темпи поширення із темпами іншими країнами та багато чого іншого. Після вибору моделі, яка найкраще описує розповсюдження і отримавши певні прогнози можна приступити до оцінки отриманих результатів. Величезна кількість різноманітних факторів впливає на розповсюдження вірусів та інфекцій. Починаючи від кількості населення, обізнаності населення, карантинних заходів соціального дистанціювання та закінчуючи кількістю зроблених тестів та якістю даних тестів. Все це впливає на якість моделі в цілому та на коректність прогнозних даних по відношенню до існуючих.

Жодна країна у світі не була готова до масштабів розповсюдження коронавірусу, але зробивши аналіз по країнах Європи можна передбачити насамперед пік захворювання, на яку дату припаде найбільша кількість інфікованих, співвідношення між інфікованими та померлими і багато чого іншого. Це дає загальне уявлення про масштаби на тенденції у розвитку пандемії.

Метою дослідження є:

- моделювання та прогнозування масштабів та наслідків поширення вірусу COVID-19 в Україні та у Європі у цілому;
- оцінювання впливу відповідних адміністративних заходів соціального дистанціювання та масового тестування на динаміку показників захворюваності.

Предметом дослідження є захворюваність, смертність, летальність та виліковність від COVID-19.

Об'єктом даного дослідження є розвиток епідемії COVID-19 в Україні та країнах Європи.

Методи дослідження включають:

- статистичний аналіз та візуалізація;
- епідеміологічні моделі SIR-SEIR-SEIRD.

Даними в роботі виступають дані інституту імені Джона Хопкінса [1] та Національної служби здоров'я України [2] в період із 01.22.2020 по 21.05.2020. Програмна реалізація була здійснена у середовищі RStudio [3] за допомогою мови програмування R.

Відповідно до мети в роботі досліджуються такі задачі:

1. Вивчення доменної галузі (поширення епідемії; фактори, які на це впливають).
2. Пошук джерел та даних стосовно поширення вірусу COVID-19.
3. Огляд традиційних методів моделювання поширення епідемії.
4. Первинна підготовка та аналіз даних.
5. Створення моделі та візуалізація даних.
6. Прогнозування та виведення підсумкових результатів.

1. АНАЛІТИЧНИЙ ОГЛЯД

1.1 Епідеміологія та процеси поширення

Епідеміологія — галузь медицини, що досліджує причини виникнення і закономірності поширення епідемій та розробляє методи боротьби з ними. Основним об’єктом вивчення є епідеміологічні процес, що виникає за наявності джерела збудника інфекції [4]. Слід зазначити, що епідеміологія і загальна медицина – це досить різні поняття. В епідеміології, смертність населення є показником поширення вірусу і є припустимим поняттям. На відміну від звичайної медицини, де смерть окремого індивідуму є поняттям недопустимим. Епідеміологія покликана для того, щоб краще зрозуміти специфіку хвороби, методи її поширення та багато чого іншого.

Зазвичай збудниками хвороби виступають віруси або бактерії. Ві́рус — група мікроорганізмів, яка відрізняється від прокаріотів і еукаріотів малими розмірами, відсутністю клітинної структури та протеїноутворювальних систем, вираженим цитотропізмом і облігатним внутрішньоклітинним паразитизмом. Є збудник багатьох інфекційних захворювань людини, тварин, рослин. Специфіку епідеміології вірусних інфекцій передусім визначають особливості внутрішньоклітинного паразитування збудників. У зовнішньому середовищі більшість вірусів знаходиться досить недовго, існування вірусних популяцій пов’язане із живими клітинами. [5]. Вірус (Covid-19) передається від людини до людини. Тому важливо зберігати соціальне дистанціювання. Експерти радять щонайменше 1.5 метри, а якщо можливо то і більше. Відстань є найголовнішим фактором у не розповсюдженні вірусу. Частинки, які людина видихає під час кашлю або при чиханні і є рознощиками хвороби. Можна сказати, що передача вірусу є контактна. Він потрапляє до організму після доторкання до очей, носу, дихальних шляхів через руки або різноманітні предмети на яких присутні заражені частинки.

Поширення вірусу, який на початковому етапі не був знайдений, є надзвичайно стрімким явищем. Якщо цей етап не проконтролювати, то хвороба

переходить на новий рівень і вже може називатися епідемією. Епідемія — масова захворюваність населення на інфекційну хворобу, що прогресує в часі та просторі в межах певного регіону і значно перевищує рівень, зареєстрований на даній території впродовж низки років. [6]. На поширення епідемії можуть впливати величезна кількість факторів. Вони можуть як позитивно вплинути на перебіг хвороби, так і погіршити стан речей.

Фактори, які впливають на поширення епідемії:

- щільність населення;
- мобільність населення (внутрішня і зовнішня міграція);
- скупченість у публічних місцях;
- порушення санітарного режиму праці і відпочинку;
- рівень санітарної культури населення;
- здійснення масової специфічної профілактики (відсутність або недостатність препаратів для масової імунізації тощо);
- незадовільна організація клінічної, лабораторної та санітарно-протиепідемічної допомоги населенню. [7]

Перелік досить масштабний і в кожному підпункті є свої власні «підводні камені». Наприклад фактор щільності заселення. Звісно, що урбанізація полегшує життя, сприяє розвитку економіки і так далі. Але саме найбільша небезпека і ховається у масових скупченнях людей. Також не менш важливим фактором є тестування на наявність вірусу в крові. За даними на 16 травня Україна посідає 47-ме місці за кількістю проведених тестів. А саме 211 614 осіб. Лідером за кількістю зроблених тестів залишаються США: там зробили 10,6 млн тестів, у Німеччині – 3,1 млн, в Італії – 2,8 млн, в Іспанії – 2,5 млн.

Відзначимо, що кількість зроблених тестів не дорівнює кількості людей, яким їх провели. Тест на коронавірус може проводитися повторно. Отже, в Україні тестують близько 4 осіб на кожну тисячу населення, у США – близько 32 осіб, у

Німеччині – близько 43 [8]. Навіть за таких умов можна сказати, що кількість тестів недостатня. Також не слід забувати про якість даних тестів. Інколи тест показує позитивний результат лише після декількох спроб. І це також негативно впливає на коректність зібраної інформації. Також прояв симптомів є не менш важливим показником. Людина може перехворіти без значних наслідків для власного здоров'я, але в той же час вона може бути розповсюджувачем вірусу.

Глибоко стурбована, якими тривожними темпами проходить розповсюдження вірусу, ВООЗ охарактеризувала COVID-19, як пандемію [9]. Тобто епідемія, яка набула світових масштабів. Чума, грип, холера, коронавірус 2019 – це наймасштабніші пандемії в історії людства, які суттєво вплинули на перебіг подій у світі.

Враховуючи величезну швидкість поширення вірусу як в межах окремого міста, так і у всьому світі, на перший план виходить оперативність та адекватність реагування державних органів на епідеміологічну ситуацію. Епідемії останніх років яскраво показали, що прогнозування та оцінка епідеміологічної ситуації далекі від ідеалу. Експерти складають прогнози на основі обмежених та недостовірних наборів даних про аналогічні віруси, а також на підставі інформації про поширення вірусу в інших країнах. Часто такі прогнози не дають навіть близького до правди результату. У світі вже давно розроблялись та успішно впроваджувались системи прогнозування та моделювання епідемій. На жаль, Україна відстає в інформатизації цієї важливої галузі науки [10]. Тож хотів би наголосити що, вірус COVID-19, це нова хвороба і навіть зараз вона залишається недослідженою на необхідному рівну, що ускладнює моделювання та прогнозування.

1.2 Огляд наукових робіт із поширення епідемій

Впродовж існування людства, епідемії або навіть пандемії неодноразово охоплювали країни або навіть континенти. За дві тисячі років налічується більш ніж 15 пандемій, і більше 300 млн смертей. Впродовж еволюції люди почали розуміти походження, що може викликати такі масштабні хвороби, а найголовніше як із ними боротися. Саме для цього і потрібно моделювати поширення хвороб, щоб знати до чого бути готовими. Проте найбільш значний прогрес у використанні методів моделювання пов’язують з появою в середині 50-х років XX століття перших електронно-обчислювальних машин (ЕОМ), та збільшення числа наукових робіт і публікацій по математичному і комп'ютерному моделюванню епідемій. У роботах того часу стали з'являтися все більш складні математичні моделі, в яких істотну роль відіграють випадкові чинники епідемічного процесу. Тому більшість моделей цього періоду мали стохастичний (імовірнісний) характер, а робочим апаратом була теорія імовірності і випадкових процесів [11]. На даний момент такий підхід стає все більш популярний, але надзвичайно складний зважаючи на декілька факторів. Лише з появою достатньої кількості даних стосовно нової хвороби можна приступати до моделювання. Модель є настільки показова, наскільки гарними є дані, що її будують. Або ж в іншому випадку дані будуть не репрезентативні. Інколи береться не перший випадок зараження, а наприклад 1000, що дає змогу більш точно побудувати модель спираючись на наявні данні. COVID - це експоненціально зростаюче захворювання, тому навіть невеликі відмінності у припущеннях можуть мати різний вплив на прогнози. Найбільше на криву впливають початкові значення. Саме вони формують вигляд кривої. Тому з появою нових значень крива буде більш відповідати фактичним даним.

Під час моделювання також робляться припущення стосовно деяких факторів. Наприклад, що люди з однаковими симптомами будуть госпіталізуватися однаково. Або що на певній території темп поширення хвороби однаковий. І такі припущення можуть, як позитивно, так і негативно вплинути на прогноз. Якщо модель дає чисельний прогноз, наприклад кількість

інфікованих або померлих, то слід розглядати це не як конкретне число, а як певний діапазон. Розуміння, що результати математичної моделі є приблизними і є ключовим фактором у правильній інтерпретації. Правильно обрана модель для побудови прогнозу, це частина головоломки, але не її вирішення.

Досить цікавий ресурс «Modeling COVID-19 Spread vs Healthcare Capacity» [12], який дає смогу підставити різноманітні початкові значення моделі та спостерігати результат. Але істотний мінус, що дана модель не прив'язується до реальних значень.

Також зустрічаються роботи в яких використовується надзвичайно ускладнена модель. На перший погляд здається, що так можна більш точно описати модель. Але надзвичайно важко підібрати змінні які корегують модель, щоб вони співпадали із фактичними даними. Інколи створюють модель із повторним зараженням. Але на даний момент неможливо сказати чи інфіковані люди, які одужали зможуть стати інфікованими знов. Така ж саме проблема і із сезонністю.

У статті «Is COVID-19 as bad as all that? Yes it probably is» [13] робляться припущення за різноманітними сценаріями. Наприклад, кількість інфікованих співпадає із фактичними значеннями або співпадає лише частково, змінюють число R, щоб побачити як це впливає на прогноз.

У статті «COVID-SIR» [14], наглядно показано, як може розвиватися поширення епідемії, зважаючи на введення карантину, соціального дистанціювання або цих двох заходів одночасно. Це дає можливість поглянути наскільки результати кардинально відрізняються.

Цікава стаття під назвою «Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy» [15] розповідаю про недоліки та переваги різних моделей на прикладі поширення вірусу в Італії. Порівнюються моделі, та обираються необхідні, зважаючи від обставин.

Цікавим рішенням неточності у прогнозі кількості інфікованих скористалася група вчених із університету Сантьяго [16]. Вони побудували не

криву, як зазвичай, а діапазон в якому кількість інфікованих може, як рости так і зменшуватися.

Також слід зазначити, що різні моделі, взявши однаковий початковий набір, роблять прогноз на різний термін. Наприклад, прогноз на день буде більш точним, але в той же час неможливо припустити коли буде пік епідемії або коли вона піде на спад. Тут вже відіграє наскільки важлива точність або довгострокова складова моделі.

Існує велика кількість моделей для прогнозування, серед яких найпопулярнішою є модель SIR. Але існує велика кількість модифікацій, які ускладнюють модель. Наприклад модель SIRS, яка враховує можливість повторного зараження або модель SEIR, яка враховує затримку. Кожна модифікація підпадає під певний тип хвороби, що допомагає у прогнозуванні.

На даний момент існують різні моделі із суттєво різними результатами. На мою думку найбільш ефективними є моделі які поєднують наявні дані (інфіковані або померлі) та експертні міркування.

1.3 SIR модель поширення епідемії

Почнемо з обрання моделі. Я обрав для свого дослідження модель SIR. Існує велика кількість епідеміологічних моделей, але серед усіх вона є, напевно, найпопулярнішою. Основна ідея моделі SIR (S – сприйнятливі, I – інфекційні, R - вилікувані) стосовно інфекційних захворювань полягає в тому, що існує три групи людей:

- S: ті, хто здоровий, але сприйнятливий до захворювання. На початку пандемії, S - це вся популяція, оскільки ніхто не має імунітету до вірусу.
- I: інфіковані люди
- R: особи, які були заражені, але з плином часу видужали або померли. Вони вже не можуть передати захворювання.

З плином часу кількість населення в кожній групі буде змінюватися за таким правилами:

1. Сприятлива група населення (S) зменшуються шляхом зараження і переходить до інфікованих (I).
2. Інфіковані (I) переходять до вилікуваних (R) шляхом одужання або смерті.

Також нам необхідні коефіцієнти, які регулюють швидкість перебігу реакцій між групами людей:

- β – швидкість зараження
- γ - швидкість відновлення

Та параметр N, що відповідає за кількість населення.

Слід зазначити, що в кожен момент часу загальна кількість населення має бути однакова:

$$S(t) + I(t) + R(t) = const \qquad (1)$$

В результаті маємо три рівняння, які описують нашу модель:

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \tag{2}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \tag{3}$$

$$\frac{dR}{dt} = \gamma I \tag{4}$$

Як візуально протікають реакції, можна переглянути на Рис. 1

Рис. 1 – Схематичне зображення SIR-моделі

Створення моделі можна поділити на два підпункти:

А) розв’язання диференціальних рівнянь

Б) знаходження оптимального значення для β, γ

В підпункті А скористаємося функцією ode, яка є функцією інтеграції за замовчуванням, із пакету deSolve. Ode – ітераційно вирішує систему звичайних диференціальних рівнянь.

Функція ode приймає як вхідні дані:

- y – початкові дані (S, I, R, N)
- times - час прогнозування
- func – модель SIR
- params – параметр β, γ.

Функція ode повертає об'єкт класу deSolve з матрицею, яка містить значення змінних стану (стовпці) у певний час. [20]

В підпункті Б скористаємося вбудованою функцією optim та модифікуємо його додавши RSS (залишкова сума квадратів):

$$RSS(\beta,\gamma) = \sum_t \left(I(t) - \hat{I}(t)\right)^2 \quad (5)$$

Завдяки цьому ми зможемо мінімізувати різницю між фактичними даними по інфікованим та прогнозними за нашою моделлю.

Модель була побудована за такими припущеннями:

1. Люди з певної групи (S, I, R) не відрізняються між собою
2. Щільність населення однакова
3. Загальна кількість населення є константою
4. Кожна людина має однаковий шанс заразитися
5. Протягом всього періоду вірус передається однаково
6. Неможливе повторне зараження
7. Хвороба розвивається з урахуванням карантинних заходів

2. ПРАКТИЧНА РЕАЛІЗАЦІЯ

2.1 Опис даних

Для побудови моделі нам необхідно мати певний об'єм інформації щодо фактичних даних розповсюдження коронавірусу. Перш за все це кількість інфікованих, вилікуваних, померлих за період часу, починаючи від першого випадку зараження. Також для візуалізації необхідно мати вік, стать або регіон чи країну. В моїй роботі я порівнюю Україну із країнами Європи. Тому необхідно мати дві таблички формату .csv, для зручності використання у програмі. Для проведення роботи була обрана мова програмування R, через зручність роботи із даними формату .csv.

Для аналізу Європи були використані щоденні дані за період із 22/01/2020 – 21/05/2020 із репозитарію інституту імені Джона Хопкінса [1]. Після форматування даних (зміна типу, видалення непотрібної інформації і тд.) була взята така інформація:

- date – дата (день, місяць)
- country – країна
- type – тип випадку (інфікований, вилікуваний, померлий)
- cases – кількість випадків на день

Для аналізу інформації стосовно України я скористався щоденними даними за період 22/01/2020 – 01/06/2020, які були зібрані Національною службою здоров’я України [2]. Порівнюючи із даними по Європі, табличка є більш інформативна та з більшою кількістю змінних.

Таблиця по Україні включає таку інформацію.:

- zvit_date – дата (день, місяць)
- priority_hosp_area - місце госпіталізації
- edrpou_hosp - код єдрпоу лікарні, в яку надійшов хворий із підозрою
- legal_entity_name_hosp – назва медичного закладу

- person_gender – стать особи, з підозрою на наявність вірусу в крові
- person_age_group - вікова група
- is_medical_worker – чи є медичним працівником
- new_confirm нові випадки в день
- active_confirm – наразі із захворюванням
- new_death – смерті за день
- new_recover – вилікувані за день

Звісно, що існують певні неточності у порівнянні моделі по Україні та країнам Європи. Перш за все, це темпи поширення вірусу. Для порівняння в Італії приблизно в 9 разів більше інфікованих ніж в Україні. Різні дати введення карантинних заходів. Також відрізняється кількість зроблених тестів за однаковий проміжок часу. Все це ускладнює порівняння моделей.

2.2 Візуалізація та початковий аналіз по Україні

В даному розділі я провів візуалізацію деяких параметрів, які нам допоможуть в майбутньому.

На Рис. 2.1 показаний розподіл хворих на COVID-19 в залежності від статі станом на 06/01/2020

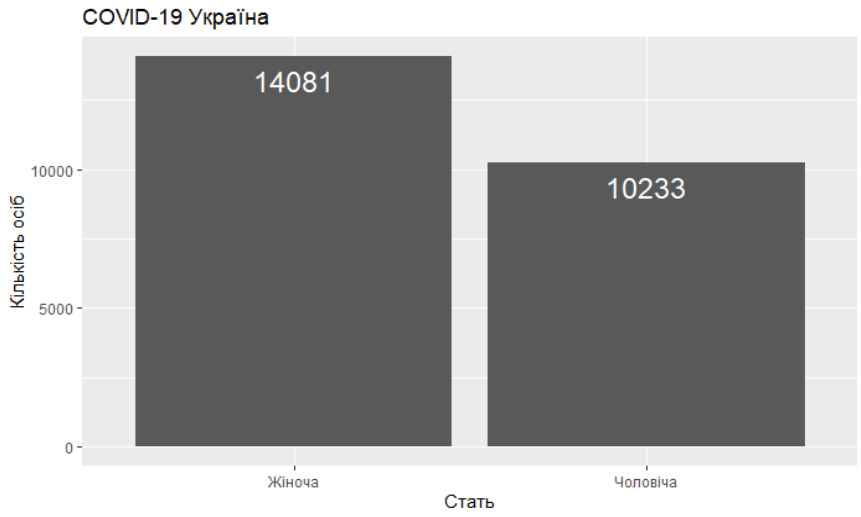

Рис. 2.1 – Розподіл за статтю осіб, що захворіли на Covid-19 станом на 06/01/2020

А тепер переглянемо скільки із цих людей є працівниками медичних закладів. (Рис. 2.2)

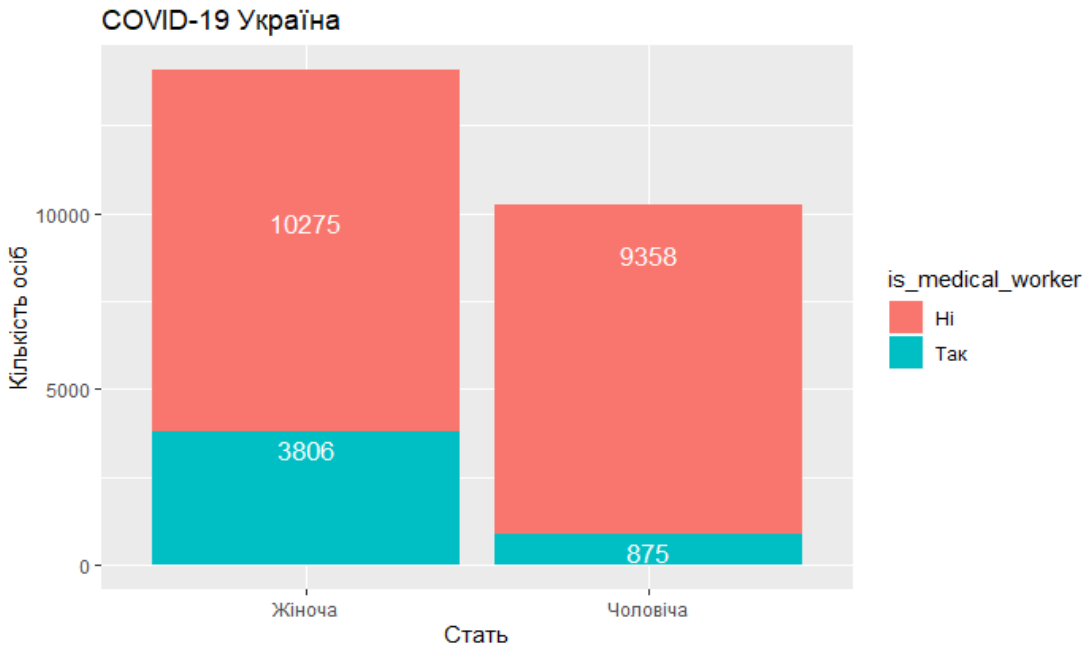

Рис. 2.2 – Розподіл за статтю осіб, що захворіли на Covid-19 станом на 06/01/2020 та приналежністю до мед. сфери

Тепер ми бачимо, що люди, які не є працівниками медичної сферою заражаються майже однаково. Тобто вірус заражає чоловіків та жінок майже однаково.

Тепер переглянемо розподіл населення за віковою групою. (Рис. 3)

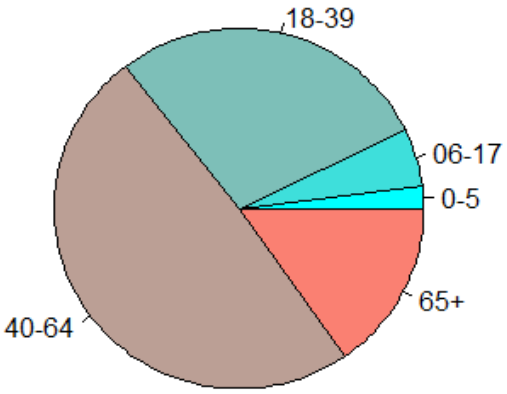

Рис. 2.3 – Розподіл за віком осіб, що захворіли на Covid-19 станом на 06/01/2020

Можна побачити, що для людей після 40 вірус є більш небезпечний. Виконуючи цю роботу впродовж тривалого проміжку часу можу сказати, що дана діаграма майже не змінюється.

Також побудуємо гладкий графік (умовно згладжує середнє значення) для кількості хворих, вилікуваних, померлих та підозрюваних на день за весь проміжок час. (Рис 4)

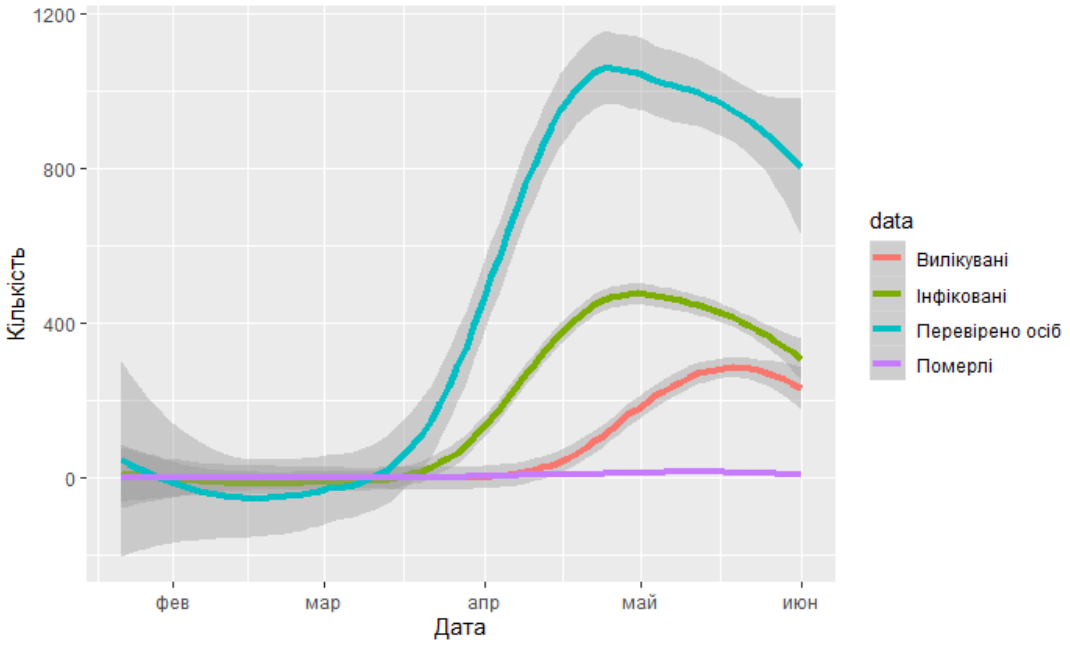

Рис. 2.4 – Статистика поширення Covid-19 по дням за увесь час станом на 06/01/2020

2.3 Побудова моделі по Україні

Виходячи із фактичних даних станом на 21/05/2020, можна почати моделювання інфікованого населення (ДОДАТОК А). Щоб модель відповідала фактичним даним скористаємося RSS для зменшення відхилення.

Були отримані такі результати (Рис 2.5):

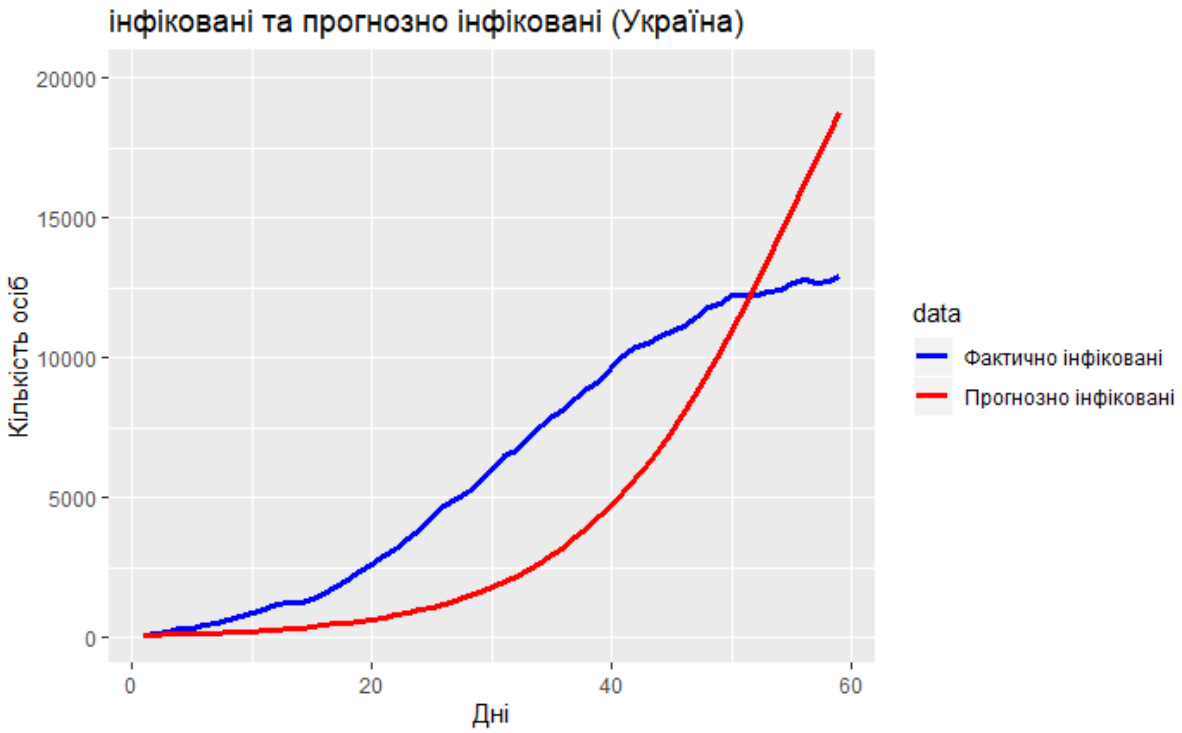

Рис 2.5 – Прогнозування кількості інфікованих на Covid-19 в період 24/03/2020 - 21/05/2020

Червона крива (Прогнозно інфіковані) завжди набуває вигляду схожого із законом нормального розподілу. Складно змінити нахил кривої, через те що впливати ми можемо тільки на коефіцієнти β та γ. Але за допомогою RSS ми мінімізуємо загальну похибку.

Для того щоб краще інтерпретувати отримані результати, скористаємося репродуктивним числом R.

$$R = \frac{\beta}{\gamma} \tag{6}$$

Це число показує кількість сприятливого населення, яке буде інфіковане при контакті із інфікованою людиною. Якщо число більше R > 1 то хвороба

поширюється. Якщо його не контролювати то хвороба може перерости в епідемію або навіть пандемію. Або якщо $R < 1$, то поширення хвороби затухає. Спираючись на дані ВООЗ [21], репродуктивне число R коливається в межах 2-2.5. Цей показник показує, що від однієї людини заражається 2-2.5 людини. Така оцінка є достатньо грубою, бо такі результати можна отримати лише в групі людей, які контактують між собою без будь-яких зовнішніх факторів. Такі, як заселеність, самоізоляція, карантинні заходи та багато чого іншого.

Виходячи із отриманих результатів по Україні, число R вийшло 1.035. Це свідчить, що епідемія досить невеликими темпами, але поширюється Кількість населення становить 37289000 [22]. При таких результатах поширення епідемії в Україні буде розвиватися так (Рис 2.6) :

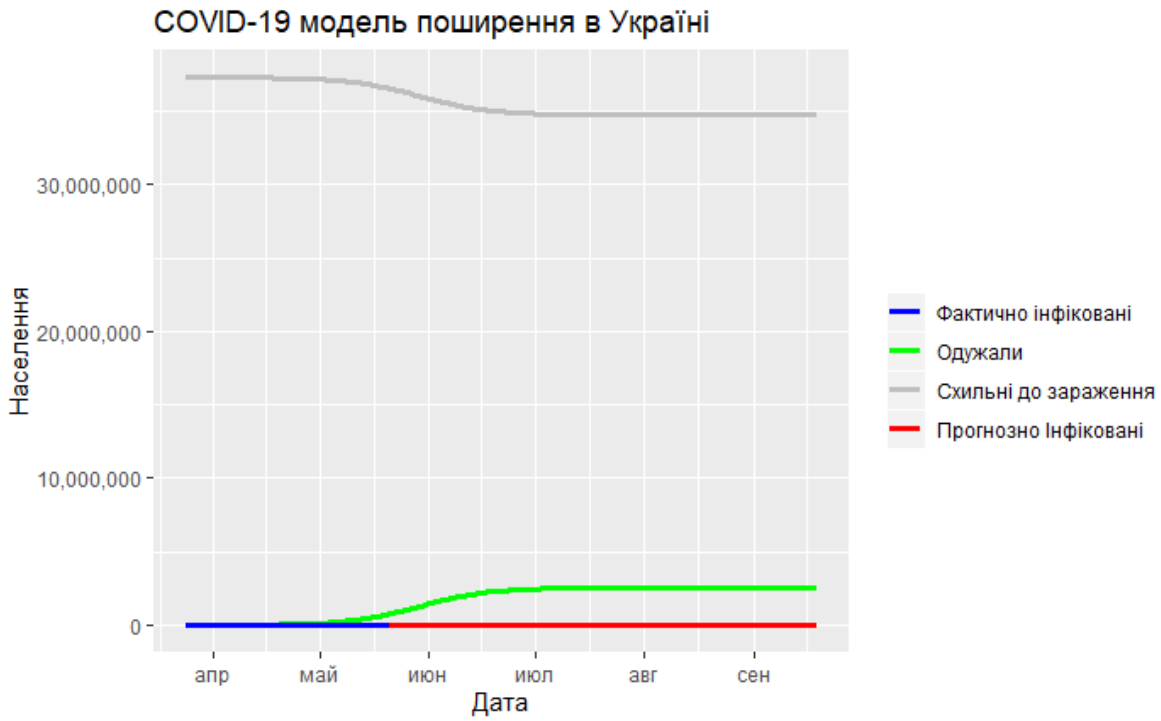

Рис. 2.6 – Модель SIR при R = 1.035 (Україна)

Збільшимо криву червоного кольору (Рис 2.7). Можна припустити, що пік епідемії припаде на кінець травня початок червня. Кількість інфікованих в пік буде близько 22.5 тис. осіб

Рис 2.7 – Кількість інфікованих за моделлю SIR при R=1.035

Тепер побудуємо модель при R = 2 [21] (Рис 2.8). Тобто, без усіляких заходів, які стримують поширення епідемії.

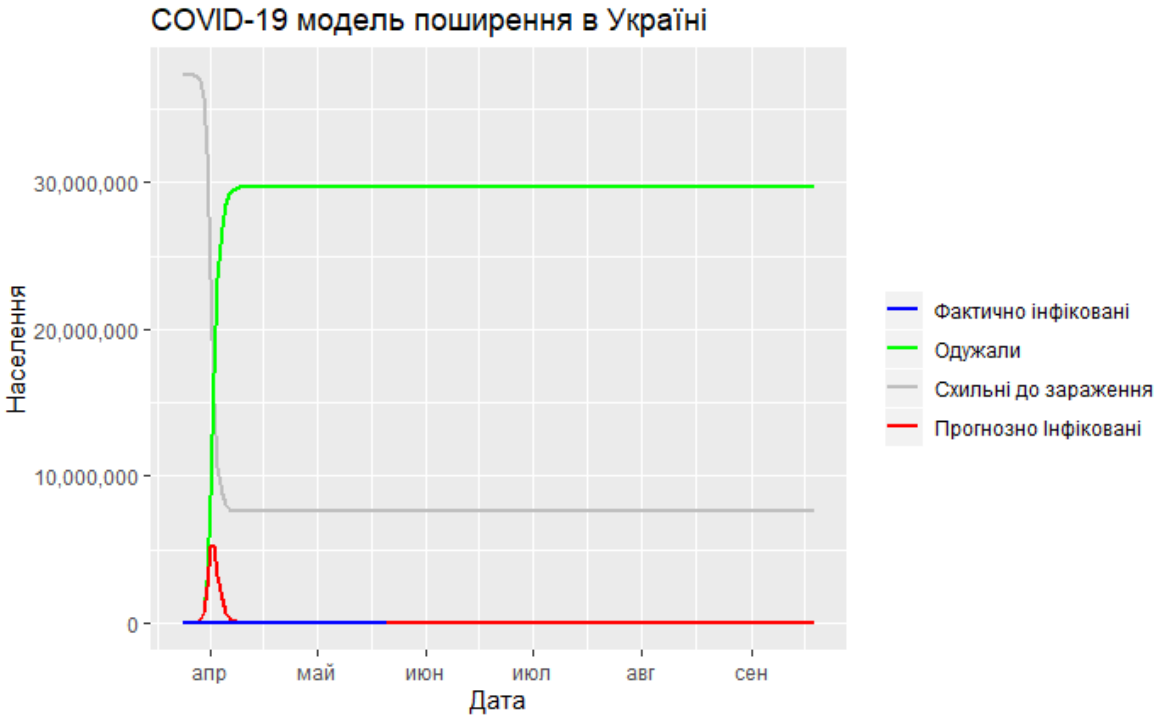

Рис 2.8 - Модель SIR при R = 2 (Україна)

Кількість інфікованих в пік епідемії за таких показників буде близько 5 млн. А кількість вилікуваних буде досягати 30 млн..

2.4 Побудова моделі по країнам Європи

Почнемо із епіцентру поширення Covid-19, а саме Італія. Після моделювання репродуктивне число R = 1.071. Для порівняння, в Україні число R вдвічі менше, що свідчить про різні темпи зараження. На Рис 2.9 можна побачити, що пік в Італії вже пройшов.

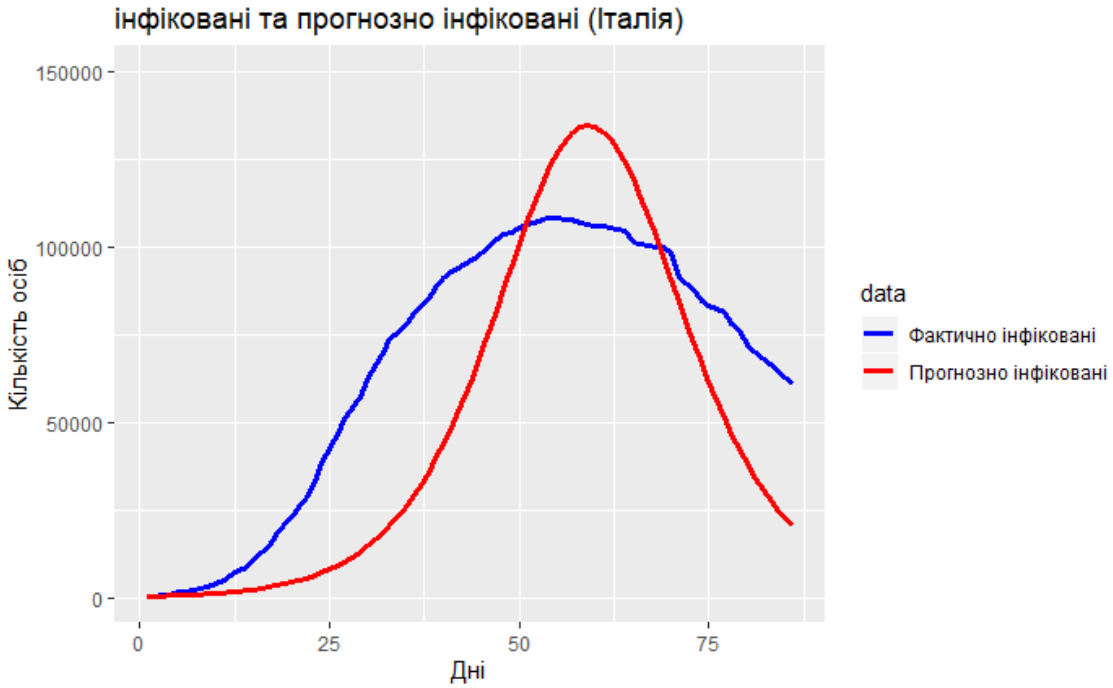

Рис 2.9 - Прогнозування кількості інфікованих в Італію в період 26/02/2020 – 22/05/2020

Побудуємо модель SIR для Італії (Рис 2.10).

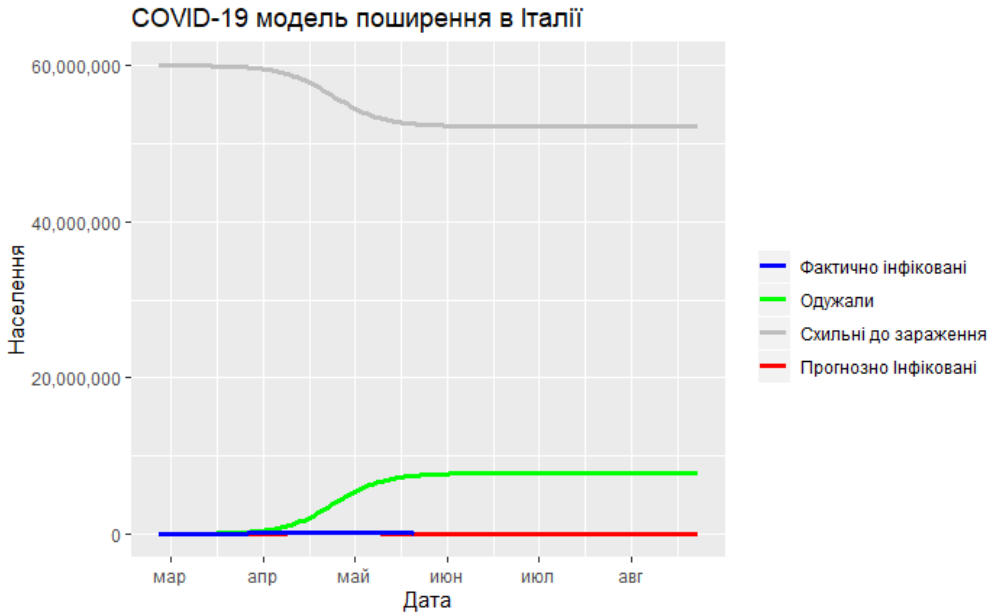

Рис. 2.10 – Модель SIR при R = 1.071 (Італія)

Кількість інфікованих в пік епідемії за таких показників буде близько 135 тис. Порівнюючи дані по Італії із даними по Україні, можна сказати, що навіть невелике відхилення числа R істотно впливає на поширення епідемії.

Побудуємо модель поширення епідемії по всім країнам Європи. Дані по населення були взяті із сайту Wikipedia [23], а саме 741 млн. Змоделюємо кількість інфікованих станом на 21/05/2020 (Рис 2.11) та станом на 14/04/2020 (Рис 2.12).

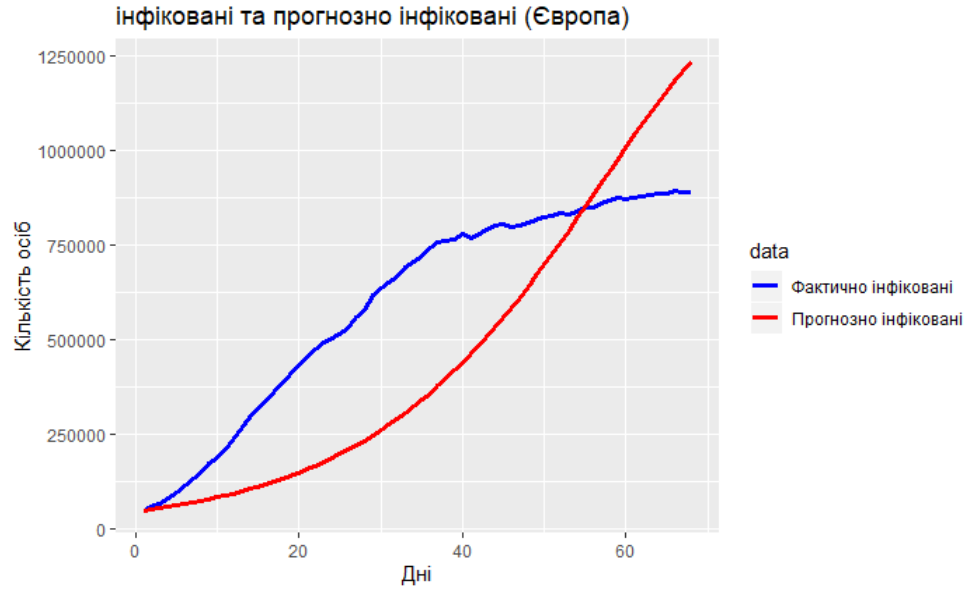

Рис 2.11 - Прогнозування кількості інфікованих в Європі станом на 21/05/2020

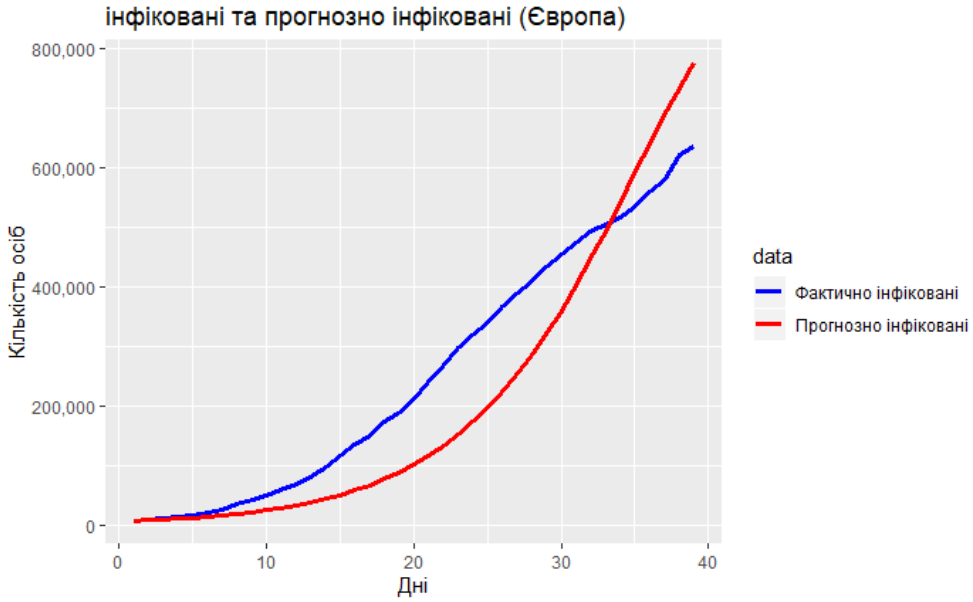

Рис 2.12 - Прогнозування кількості інфікованих в Європі станом на 14/04/2020

Порівнюючи із більш ранньою версією кількість інфікованих почала зменшуватися, що позитивно впливає з точки зору поширення епідемії, але негативно з точки зору моделі.

Побудуємо модель поширення вірусу із отриманим числом R = 1.063 (Рис 2.13). Таке число є логічним, бо епідемія в Європі розвивається швидше ніж в Україні (R = 1.035), але повільніше ніж в Італії (R = 1.071).

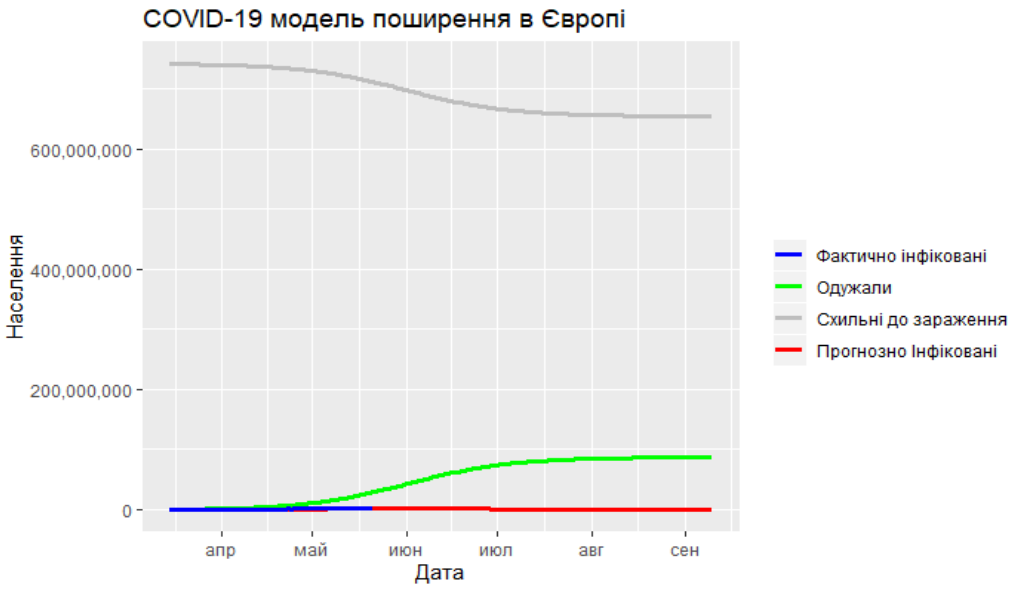

Рис. 2.13 – Модель SIR при R = 1.063 (Європа)

Збільшимо червону, зелену та сині криві (Рис 2.14). Кількість інфікованих за таким прогнозом в пік епідемії виходить 1 370 тис., що в порівнянні із загальним населенням Європи є невеликим числом .

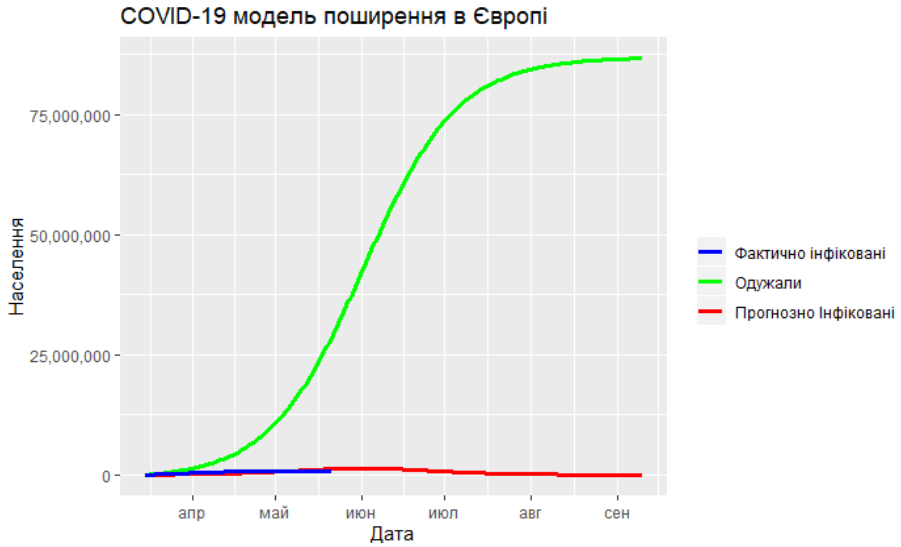

Рис 2.14 - Модель SIR при R = 1.063 (Європа)

Для оцінки точності, скористаємося MAE (Середня абсолютна помилка)

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y}_j|$$

MAE вимірює середню величину помилок у наборі прогнозів. Це середнє значення для тестової вибірки абсолютних різниць між прогнозуванням і фактичним спостереженням. [24]

Також скористаємося MAPE (Середня абсолютна помилка у відсотках)

$$MAPE = \frac{1}{n}\sum_{j=1}^{n} \left|\frac{y_j - \hat{y}_j}{y_i}\right|$$

Для моделі по Україні MAE становить 2601, MAPE дорівнює 0.52. В Італії MAE становить 23 390, MAPE дорівнює 0.45. В країнах Європи MAE - 213 358, MAPE – 0.39. Слід зазначити, що це лише припущення, але навіть така помилка є допустимою у порівнянні із загальною кількістю інфікованих та моделями інших дослідників.

Для прикладу по Україні візьмемо модель Київської школи економіки. За даними моделі, загальна кількість випадків, станом на 04/06/2020 складає близько 100,000 осіб vs 25,000 офіційно підтверджених. [25] Слід зазначити що, дана стаття з певною періодичністю оновлюється. Тому на отриманих даних робиться припущення, що сезонність або відсутня, або помірна. Але також зазначається, що сезонність в різних країнах та регіонах протікає по різному. Точність моделі по відношенню за інфікованими складає 25%.

В Італії для порівняння візьмемо модель університету Туріна [26]. За фактичними даними станом на початок квітня, прогноз стосовно кількості інфікованих на початок травня був у діапазоні від 50 тис. до 140 тис.(в залежності від різних параметрів). Дивлячись на фактичні дані на початок травня, а саме близько 100 тис., можна сказати, що за певних параметрів модель буде близькою до реальної ситуації, але перевірити це можливо лише з часом.

ВИСНОВКИ

В даній роботі було проведене дослідження стосовно епідеміологічних моделей. Були проаналізовані джерела в яких описується дана тематика. Була розглянута модель SIR для побудови кривих поширення вірусу.

Була зібрана інформація, яка дає смогу проаналізувати темпи поширення епідемії на території України та Європи. Після цього було проведено підготовка та упорядкування та відокремлення потрібних даних. Після проведення первинного аналізу була зроблена візуалізація даних по Україні.

В даній роботі були побудовані моделі з різними значеннями β та γ, які описуюсь розповсюдження коронавірусу. Було розглянуто, як змінюється кількість інфікованих в залежності від початкових параметрів. Був зроблений прогноз кількості інфікованого населення і за певних параметрів він є подібний до реальних даних. Всі побудовані моделі є лише припущеннями, тому їх не варто сприймати як ідеальну модель, яка повністю відповідає дійсності. Звісно, що ця модель досить спрощена і вона не враховує всі фактори, які впливають на розповсюдження. Наприклад, зменшення спроможності медичної допомого через відсутність належної кількості тестів або послаблення чи навпаки посилення карантинних заходів. Але навіть при візуальному порівняні даних по Україні на Європі можна зробити висновки, що моделювання та отриманий прогноз лише допомагає у боротьбі із вірусом. Це дослідження допомагає оцінити масштаб поширення, та можливий сценарій за яким буде розвиватися хвороба. Це лише частина головоломки, яка можливо допоможе вирішити більш глобальну головоломку.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE : Data. GitHub. URL: https://github.com/CSSEGISandData/COVID-19 (Last accessed: 09.06.2020).
2. Національна служба здоров'я України : вебсайт. URL: https://nszu.gov.ua/covid/dashboard (дата звернення: 09.06.2020).
3. RStudio : вебсайт. URL: https://rstudio.com/ (дата звернення: 09.06.2020).
4. Епідеміологія. *Енциклопедія сучасної України* : вебсайт. URL: http://esu.com.ua/search_articles.php?id=17930 (дата звернення: 09.06.2020).
5. Віруси. *Фармацевтична енциклопедія* : вебсайт. URL: https://www.pharmencyclopedia.com.ua/article/1764/virusi (дата звернення: 09.06.2020).
6. Епідемія. *Енциклопедія сучасної України* : вебсайт. URL: http://esu.com.ua/search_articles.php?id=17931 (дата звернення: 09.06.2020).
7. Основне про пандемії [Електронний ресурс] – режим доступу: http://esu.com.ua/search_articles.php?id=17931
8. Пандемія коронавірусу: скільки тестів провели в Україні та інших країнах світу. *Слово і діло* : аналітичний портал. URL: https://www.slovoidilo.ua/2020/05/16/infografika/svit/pandemiya-koronavirusu-skilky-testiv-provely-ukrayini-ta-inshyx-krayinax-svitu (дата звернення: 09.06.2020).
9. WHO Timeline – COVID-19. *World Health Organization* : website. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 (Last accessed: 09.06.2020).
10. Пшеничний О. Ю., Чорней І. М., Шаховська Н. Б., Литвин В. В. Аналіз сучасних програмних засобів моделювання поширення вірусних

захворювань. *Вісник Національного університету "Львівська політехніка" : Інформаційні системи та мережі*. 2010. Вип. 673. С. 154–162.

11. Котвіцька А. А., Суріков О. О. Науково-практичні підходи до моделювання епідемічних процесів та фармацевтичного забезпечення. *Соціальна фармація: стан, проблеми та перспективи* : міжнар. наук.-практ. інтернет-конф., 17–20 берез. 2014 р. Х. : Вид-во НФаУ, 2014. С. 110–116. URL: https://socpharm.nuph.edu.ua/files/2014/03/%D0%9A%D0%BE%D1%82%D0%B2%D1%96%D1%86%D1%8C%D0%BA%D0%B0-%D0%90.-%D0%90.-%D0%A1%D1%83%D1%80%D1%96%D0%BA%D0%BE%D0%B2-%D0%9E.-%D0%9E.pdf (дата звернення: 09.06.2020).
12. Modeling COVID-19 Spread vs Healthcare Capacity. URL: https://alhill.shinyapps.io/COVID19seir/ (Last accessed: 09.06.2020). License: CC BY-SA 4.0.
13. Francis. Is COVID-19 as bad as all that? Yes it probably is. *Econometrics By Simulation*. URL: http://www.econometricsbysimulation.com/2020/04/is-covid-19-as-bad-as-all-that-yes-it.html (Last accessed: 09.06.2020).
14. Verardi V. COVID – SIR. *Learning from the curve* : website. URL: https://www.learningfromthecurve.net/epidemic-models/2020/04/13/covid-sir (Last accessed: 09.06.2020).
15. Giordano G., Blanchini F., Bruno R., Colaneri P., Di Filippo A., Di Matteo A., Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *naturemedicine*. URL: https://www.nature.com/articles/s41591-020-0883-7#code-availability (Last accessed: 09.06.2020).
16. Covid-19 prediction / Created by : M. Oviedo, M. Febrero. URL: http://modestya.securized.net/covid19prediction/ (Last accessed: 09.06.2020).

17. Goedecke D. M., Feng Yu, Bobashev G., Epstein J. M., Morris R. J. Stochastic Equation-Based Model of a Global Epidemic : Research Triangle Institute International, RTI Project No. 0209149.003.003. North Carolina. 2007.

18. Плавинский С. Л. Моделирование ВИЧ-инфекции и других заразных заболеваний человека и оценка численности групп риска. Введение в математическую эпидемиологию. М. : Акварель, 2010. 99 с.

19. Кузнецов Ю. А., Мичасова О. В. Теоретические основы имитационного и компьютерного моделирования экономических систем : учеб.-метод. матер. Нижний Новгород, 2007. 112 с.

20. Package deSolve: Solving Initial Value Differential Equations in R. URL: https://cran.r-project.org/web/packages/deSolve/vignettes/deSolve.pdf (Last accessed: 09.06.2020).

21. Coronavirus disease 2019 (COVID-19) Situation Report – 46 : Data as reported by national authorities by 10AM CET 06 March 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_2 (Last accessed: 09.06.2020).

22. Оцінка чисельності наявного населення України. *Telegram*. URL: https://t.me/dmytro_dubilet/578 (дата звернення: 09.06.2020).

23. Європа. *Вікіпедія* : вільна енциклопедія. URL: https://uk.wikipedia.org/wiki/%D0%84%D0%B2%D1%80%D0%BE%D0%BF%D0%B0 (дата звернення: 09.06.2020).

24. MAE and RMSE — Which Metric is Better? URL: https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d (Last accessed: 09.06.2020)

25. Covid-19 моделювання поширення захворювання URL:https://kse.ua/wp-content/uploads/2020/06/KSE-Institute_Covid-19_Modeling_June-event_vFin.pdf (Last accessed: 09.06.2020)

26. Alberto Godio1, Francesca Pace, Andrea Vergnano SEIR Modeling of Italian Epidemic of SARS-CoV-2 URL: https://www.preprints.org/manuscript/202004.0073/v1 (Last accessed: 09.06.2020)

ДОДАТОК А

```
knitr::opts_chunk$set(echo = TRUE)
library(skimr)
library(ggplot2)
library(dplyr)
library(tidyr)
library(tidyverse)
library(lubridate)
library(Metrics)



#зчитування данних
getwd()
ukraine <- read.csv("dataset_04.csv", sep = ";", header = TRUE)

str(ukraine)
head(ukraine)

#перевірка на належності до класу
class(ukraine$edrpou_hosp)

#зміна типу
ukraine$legal_entity_name_hosp = as.character(ukraine$legal_entity_name_hosp)
ukraine$zvit_date = as.Date(ukraine$zvit_date, "%d.%m.%Y")
ukraine = ukraine[order(ukraine$zvit_date),]

str(ukraine)
summary(ukraine)

# унікаьлні значення в колонці
nlevels(ukraine$zvit_date)
levels(ukraine$person_gender)
unique(ukraine$zvit_date)

data = ukraine
data = data %>% select(zvit_date, person_gender, new_confirm) %>% group_by(person_gender)

number_of_cases = summarise(data, number = sum(new_confirm))
number_of_cases
number_of_cases = number_of_cases[c(1,3),]
which(number_of_cases$person_gender == "Жін.")
ggplot(number_of_cases, aes(person_gender, number)) + geom_bar(stat = "summary") + geom_text(aes(label=number), vjust=1.6, color="white", size=6)+labs(y = "Кількість осіб", x = "Стать", title = "COVID-19 Україна")
```

```
data = ukraine
data = data %>% select(zvit_date, person_gender, new_confirm, is_medical_worker) %>% group_by(person_gender, is_medical_worker)

number_of_cases = summarise(data, number = sum(new_confirm))
number_of_cases
number_of_cases = number_of_cases[c(1,2,5,6),]
which(number_of_cases$person_gender == "Жін.")
ggplot(number_of_cases, aes(person_gender, number, fill = is_medical_worker)) + geom_bar(stat = "summary") + geom_text(aes(label=number), vjust=1.4, color="white", size=4.5)+
  labs(y = "Кількість осіб", x = "Стать", title = "COVID-19 Україна")

data = ukraine
data = data %>% select(person_gender, new_confirm, person_age_group) %>% group_by(person_age_group)

number_of_cases = summarise(data, number = sum(new_confirm))
number_of_cases
number_of_cases = number_of_cases[c(1,2,3,4,5),]



palet=colorRampPalette(c("cyan1", "salmon"))
colors=palet(5)

pie(number_of_cases$number, number_of_cases$person_age_group, col = colors)

all_confirmed = ukraine %>% select(zvit_date, new_confirm, new_susp) %>% group_by(zvit_date) %>%
  summarise(total_confirm = sum(new_confirm), total_cases = sum(new_susp))

df = tbl_df(data.frame(all_confirmed))
df

df %>% ggplot(aes(x = zvit_date, y = total_cases))+ geom_line() + geom_smooth(color = "black") + geom_line( y = df$total_confirm) + geom_smooth( aes(y = df$total_confirm))

suspect = ukraine %>% select(zvit_date, new_susp) %>% group_by(zvit_date) %>%
  summarise(total = sum(new_susp))
all_confirm = ukraine %>% select(zvit_date, new_confirm) %>% group_by(zvit_date) %>%
  summarise(total = sum(new_confirm))
all_dead = ukraine %>% select(zvit_date, new_death) %>% group_by(zvit_date) %>%
  summarise(total = sum(new_death))
```

```
recover = ukraine %>% select(zvit_date, new_recover) %>%
group_by(zvit_date) %>%
  summarise(total = sum(new_recover))

suspect$data <- 'Перевірено осіб'
all_confirm$data <- 'Інфіковані'
all_dead$data <- 'Померлі'
recover$data <- 'Вилікувані'

df <- rbind.data.frame(suspect, all_confirm, all_dead, recover)

ggplot(df, aes(x = zvit_date, y = total))+
  geom_smooth(aes(colour = data), size = 1.7, span = 0.54)+
  labs(
    y = "Кількість",
    x = "Дата",
    title = "COVID-19 Україна")

SIR <- function(time, state, parameters) {
  par <- as.list(c(state, parameters))
  with(par, {
    dS <- -beta * I * S / N
    dI <- beta * I * S / N - gamma * I
    dR <- gamma * I
    list(c(dS, dI, dR))
  })
}

coronavirus <- read.csv("SIR23_05.csv", sep = ",", header = TRUE)

df <- coronavirus %>%
  filter(country == "Ukraine") %>%
  group_by(date, type) %>%
  summarise(total = sum(cases, na.rm = TRUE)) %>%
  pivot_wider(
    names_from = type,
    values_from = total
  ) %>%
  arrange(date) %>%
  ungroup() %>%
  mutate(active = confirmed - death - recovered) %>%
  mutate(
    confirmed_cum = cumsum(confirmed),
    death_cum = cumsum(death),
    recovered_cum = cumsum(recovered),
    active_cum = cumsum(active)
```

```
  )

library(tidyverse)
library(lubridate)



df$date = as.Date(df$date, "%Y-%m-%d")


Infected <- subset(df, date >= ymd("2020-03-24") & date <= ymd("2020-05-21"))$active_cum
Infected
Day <- 1:(length(Infected))

N <- 3728900
init <- c(
  S = N - Infected[1],
  I = Infected[1],
  R = 1
)

RSS <- function(parameters) {
  names(parameters) <- c("beta", "gamma")
  out <- ode(y = init, times = Day, func = SIR, parms = parameters)
  fit <- out[, 3]
  sum((Infected - fit)^2)
}



library(deSolve)
Opt <- optim(c(0.5, 0.5),
             RSS,
             method = "L-BFGS-B",
             lower = c(0, 0),
             upper = c(3, 3)
)
Opt$message
```


```{r}
Opt_par <- setNames(Opt$par, c("beta", "gamma"))

#Opt_par[1] =
#Opt_par[2] =
Opt_par

R0 <- as.numeric(Opt_par[1] / Opt_par[2])
```

```
R0

sir_start_date <- "2020-03-24"
t <- 1:as.integer(ymd("2020-05-22") - ymd(sir_start_date))

fitted_cumulative_incidence <- data.frame(ode(
  y = init, times = t,
  func = SIR, parms = Opt_par
))

library(dplyr)
fitted_cumulative_incidence <- fitted_cumulative_incidence %>%
  mutate(
    Date = ymd(sir_start_date) + days(t - 1),
    Country = "Ukraine",
    cumulative_incident_cases = Infected
  )

true_infected = fitted_cumulative_incidence %>% select(time, I) %>% group_by(time) %>%
  summarise(total = sum(I))
predicted_infected = fitted_cumulative_incidence %>% select(time, cumulative_incident_cases) %>% group_by(time) %>%
  summarise(total = sum(cumulative_incident_cases))

true_infected$data <- 'Фактично Інфіковані'
predicted_infected$data <- 'Прогнозно інфіковані'

x = mae(Infected, fitted_cumulative_incidence$I)
x

y = rmse(Infected, fitted_cumulative_incidence$I)
y

df <- rbind.data.frame(true_infected, predicted_infected)

ggplot(df, aes(x = time, y = total)) +
  geom_line(aes(color = data), size = 1.4) +
  labs(y = "Кількість осіб", x = "Дні", title = "інфіковані та прогнозно інфіковані (Україна)")+
  ylim(93, 20000)+
  scale_colour_manual(values = c('Фактично Інфіковані' = 'red', 'Прогнозно інфіковані' = 'blue'), labels = c("Фактично інфіковані", "Прогнозно інфіковані") )

Opt_par
```

```
R0 <- as.numeric(Opt_par[1] / Opt_par[2])
R0

t <- 1:180

fitted_cumulative_incidence <- data.frame(ode(
  y = init, times = t,
  func = SIR, parms = Opt_par
))

fitted_cumulative_incidence <- fitted_cumulative_incidence %>%
  mutate(
    Date = ymd(sir_start_date) + days(t - 1),
    Country = "Ukraine",
    cumulative_incident_cases = I
  )
fitted_cumulative_incidence
head(fitted_cumulative_incidence)



fitted_cumulative_incidence %>%
  ggplot(aes(x = Date)) +
  geom_line(aes(y = I, colour = "red"), size = 1.2) +
  geom_line(aes(y = S, colour = "grey"), size = 1.2) +
  geom_line(aes(y = R, colour = "green"), size = 1.2) +
  geom_line(aes(y = c(Infected, rep(NA, length(t) - length(Infected))),
colour = "blue"), size = 1.2) +
  scale_y_continuous(labels = scales::comma) +
  labs(
    y = "Населення",x = "Дата", title = "COVID-19 модель поширення в
Україні") +
  scale_colour_manual(
    name = "",
    values = c(red = "red", grey = "grey", green = "green", blue = "blue"),
    labels = c("Фактично інфіковані", "Одужали","Схильні до зараження",
"Прогнозно Інфіковані")
  )


fitted_cumulative_incidence %>%
  ggplot(aes(x = Date)) +
  geom_line(aes(y = I, colour = "red")) +
  geom_line(aes(y = S, colour = "black")) +
  geom_line(aes(y = R, colour = "green")) +
  geom_line(aes(y = c(Infected, rep(NA, length(t) - length(Infected))),
colour = "blue")) +
  scale_y_log10(labels = scales::comma) +
  labs(
```

```
    y = "Населення",x = "Дата", title = "COVID-19 модель поширення в Укаїні") +
  scale_colour_manual(
    name = "",
    values = c(red = "red", black = "grey", green = "green", blue = "blue"),
    labels = c("Схильні до зараження", "Фактично інфіковані","Одужали", "Прогнозно Інфіковані")
  )


fitted_cumulative_incidence %>%
  ggplot(aes(x = Date)) +
  geom_line(aes(y = I), colour = "red", size = 1.3) +
  scale_y_continuous(labels = scales::comma) +
  scale_colour_manual(
    name = "",
    values = c(red = "red", black = "grey", green = "green", blue = "blue"),
    labels = c("Схильні до зараження", "Фактично інфіковані", "Одужали", "Інфіковані")) +
  labs(y = "Населення",x = "Дата", title = "COVID-19 Прогнозно інфіковані в Україні")+
  scale_y_continuous(labels = scales::comma)


fit <- fitted_cumulative_incidence


#пік пандемії
fit[fit$I == max(fit$I), c("Date", "I")]
```