

13. Registry of Open Data on AWS. URL: <https://registry.opendata.aws/> (date of access: 20.11.2020).

14. The Netflix Prize. URL: <https://www.netflixprize.com/> (date of access: 20.11.2020).

15. Big Data School від Київстар. URL: <https://bigdata.kyivstar.ua/school/> (дата звернення: 20.11.2020).

ДО МОДЕЛЮВАННЯ ЕПІДЕМІЇ COVID-19 В УКРАЇНІ

Сіницький Микола Євгенович,

кандидат фізико-математичних наук, доцент,
доцент кафедри економіко-математичних дисциплін
та інформаційних технологій,
Національна академія статистики, обліку та аудиту

Епідемія COVID-19 завдає суттєвої шкоди народу України та всьому світові. Тому моделювання її розвитку для побудови прогнозів протидії їй є одним із найважливіших завдань уряду України та всієї вітчизняної статистичної науки.

Слід зауважити, що в цьому напрямі багато чого зроблено, але є й суттєві прогалини. На початку епідемії в Україні на основі розпорядження Кабміну України від 03.02.2020 № 93-р та за розпорядженням Президії НАН України від 03.02.2020 № 198 було створено робочу групу з математичного моделювання проблем, обумовлених COVID-19. До складу групи залучили провідних науковців Інституту проблем математичних машин і систем (базова установа), Інституту економіки та прогнозування, Інституту кібернетики ім. В.М. Глушкова, Інституту математики, Інституту теоретичної фізики НАН України, КНУ імені Тараса Шевченка, ДУ «Інститут епідеміології та інфекційних хвороб ім. Л.В. Громашевського» – загалом десять докторів і п'ять кандидатів переважно фізико-математичних наук.

Ця група на момент виходу першого публічного повідомлення від імені її керівника [1] зробила розрахунки короткострокових прогнозів щодо динаміки кількості захворювань на COVID-19 в Україні виходячи з наявної на I квартал 2020 року офіційної статистики про перебіг хвороби.

Для прогнозування було використано дві добре відомі математичні моделі: *SIR* (*Susceptible-Infected-Removed model*) та *SEIR* (*Susceptible-Exposed-Infected-Removed model*) [2].

Перша модель описує динаміку хвороб з набуттям сталого імунітету і є базовою, друга – її модифікація. Якщо за *SIR*-моделлю кожен із індивідумів, що входять в досліджувану популяцію, може перебувати в одному з трьох можливих станів: *Susceptible* (потенційно сприйнятливий до зараження), *Infected* (Інфікований), *Removed* (одужав і нечутливий до зараження) і передбачається, що індивідуми можуть стати несприйнятливими до зараження тільки після повного одужання, тобто в якісь моменти часу для кожного

можлива послідовність переходів *Susceptible* → *Infected* → *Removed*, то в *SEIR*-моделі додано ще один можливий стан – *Exposed* (заражений, що знаходиться в інкубаційному періоді), тобто враховано наявність інкубаційного періоду в хвороби.

Обидві моделі представляють системи диференціальних рівнянь першого порядку зі сталими коефіцієнтами та заданими початковими умовами.

SIR-модель:

$$\frac{dS}{dt} = \frac{-\beta \cdot S \cdot I}{N}; \frac{dI}{dt} = \frac{\beta \cdot S \cdot I}{N} - \gamma I; \frac{dRe}{dt} = \gamma I \quad ; \quad (1), (2), (3)$$

$$S(0) = S_0; I(0) = I_0; \Re(0) = \Re_0 \quad (4)$$

У *SEIR*-моделі замість рівняння (2) включено

$$\frac{dE}{dt} = \frac{\beta \cdot S \cdot I}{N} - \delta E; \frac{dI}{dt} = \delta E - \gamma I; E(0) = E_0, \quad (5), (6), (7)$$

де S – чисельність здорових, але сприйнятливих до хвороби осіб у момент часу t ; I – чисельність інфікованих осіб у момент часу t ; Re – чисельність вибулих (таких, що одужали, і померлих) осіб у момент часу t ; N – чисельність популяції, охопленої епідемією; β – коефіцієнт інтенсивності контактів осіб у момент часу t ; γ – коефіцієнт інтенсивності одужання осіб у момент часу t ; δ – коефіцієнт інтенсивності затримки прояву хвороби; E – чисельність хворих осіб в інкубаційному періоді.

Коефіцієнт β представляє відношення двох параметрів:

$$\beta = \frac{R_0}{T_{inf}}, \quad (8)$$

де R_0 – коефіцієнт репродукції, тобто середня кількість заражень, які спричиняє одна особа; T_{inf} – деяка характеристика саме вірусу та організму людини.

Коефіцієнти γ і δ мають таку структуру:

$$\gamma = \frac{1}{T_{inf}}; \quad \delta = \frac{1}{T_{inc}}, \quad (9), (10)$$

де T_{inc} – тривалість інкубаційного періоду.

Описані системи диференціальних рівнянь є нелінійними, проте обмеження моделей полегшують пошук їх часткових розв'язків. Так, для *SIR*-моделі має виконуватися:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dRe}{dt} = 0 \quad (11)$$

звідки

$$S(t) + I(t) + Re(t) = Const = N \quad (12)$$

Завдяки (12) розв'язувати систему (1)–(3) можна тільки для двох змінних. Крім того, оскільки

$$\frac{\beta}{\gamma} = R_0, \quad (13)$$

тоді рівняння (2) можна переписати як:

$$\frac{dI}{dt} = \left(\frac{R_0 S}{N} - 1\right) \gamma I, \quad (14)$$

звідки слідує, що в разі $R_0 > \frac{N}{S(0)}$, настає $\frac{dI}{dt}(0) > 0$, і відбувається спалах епідемії, та навпаки – за $R_0 < \frac{N}{S(0)}$ епідемія не починається.

Інтегрування після ділення рівняння (1) на рівняння (3) і відокремлення змінних дає

$$S(t) = S(0) e^{-R_0(\Re(t) - \Re(0))/N} \quad (15)$$

звідки слідує, що число вибуху в кінці епідемії ($t \rightarrow \infty$) дорівнює

$$\Re(\infty) = N - S(0) e^{-R_0(\Re(t) - \Re(0))/N} \quad (16)$$

Параметри β і γ (R_0 і T_{inf}) відіграють вирішальну роль у прогнозуванні динаміки пандемії за *SIR*-моделлю. Все залежить від коректності визначення їхніх значень. В Інтернеті (див., напр. [3]) можна побачити результати *SIR*-моделювання у вигляді *gif*-анімацій та оцінити чутливість прогнозу до змін β і γ і порівняти з наявною ситуацією.

В роботі [1] наведено прогноз, побудований на оцінках Оксфордської групи дослідників: $R_0 = 2,25$ і $T_{inf} = 4,5$. Згідно з ним максимум захворілих мав спостерігатися в середині травня 2020 року і сягати рівня $2 \cdot 10^6$ осіб. Реальну ситуацію можна побачити на графіку, показаному на рис. 1. З нього видно, що реальна кількість захворілих на 1,5–2 порядки менша за прогнозовану, і схоже на те, що ніякої першої хвилі не було, а Україна знаходиться ще на початку

пандемії. На це вказує те, що всі залежності, представлені на рис. 1, добре описуються експоненціальною залежністю виду $y = a \cdot \exp(b \cdot x)$ (рис. 2), що характерно для початкової, висхідної гілки логістичної функції.

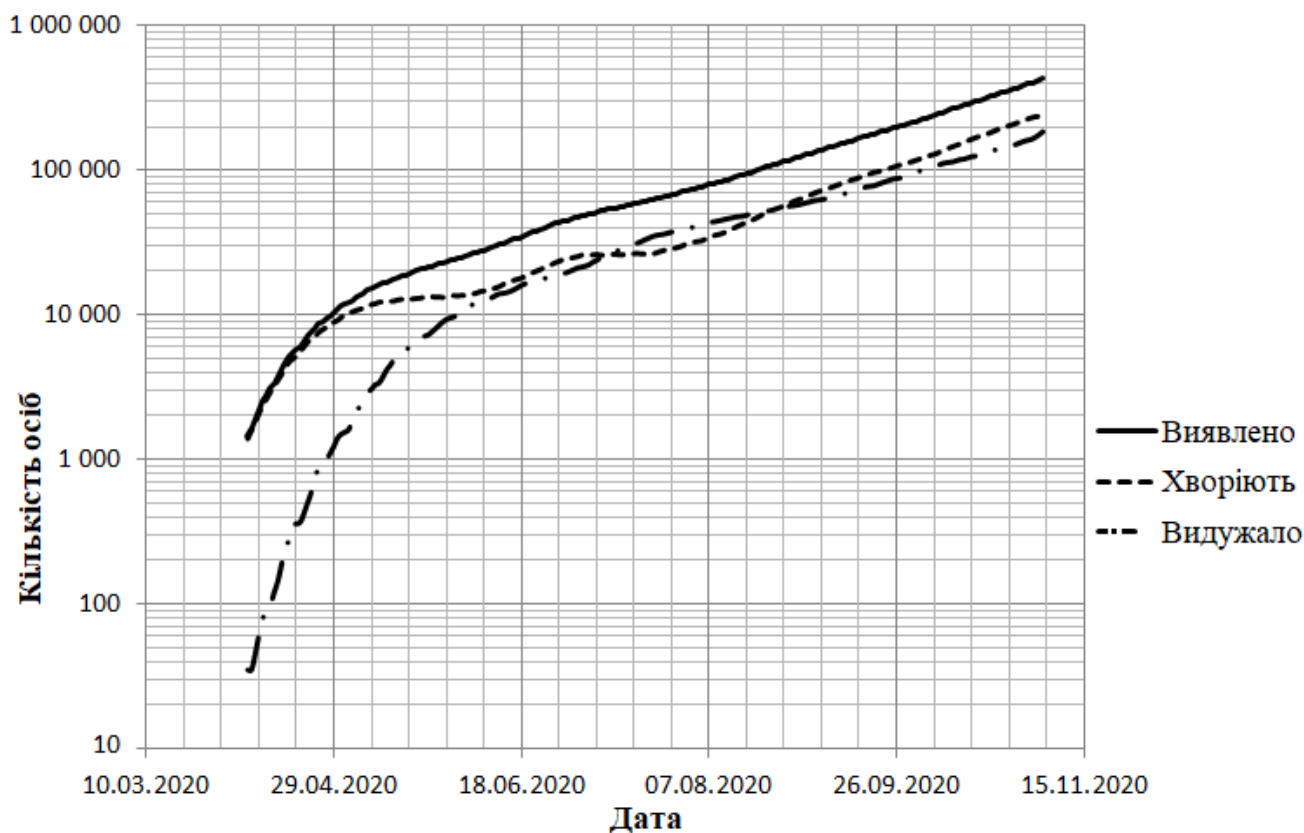


Рис. 1. Динаміка захворювання на COVID-19 в Україні
Джерело: [4]

Звернемо увагу на те, що статистичні дані [4] містять інформацію, достатню для оцінювання коефіцієнта γ в рівнянні (3) шляхом побудови лінійної регресії показника DRe (дорівнює щодобовій чисельності вибулих осіб, яку можна вважати оцінкою першої похідної показника вибування) на кількість інфікованих на той самий момент часу (добу від початку спостереження). Отримані результати показані на рис. 4. Лінійна регресія має інтерцепт (коефіцієнт $a = -8,18134$), але він визнаний статистично незначущим ($p = 0,89$) і тому таким, що не відрізняється від нуля. Звідси оцінка $\hat{\gamma}$ дорівнює коефіцієнту $\hat{\gamma} = b = 0,00872$.

В аналогічний спосіб оцінюємо величину γ у $\left(\frac{\beta \cdot S}{N} - \gamma\right)$ рівнянні (2) (рис. 5). Віднявши зі значення знайденого коефіцієнта b оцінку $\hat{\gamma}$, отримуємо оцінку величини $\hat{\gamma}$. Далі з урахуванням (13) знаходимо $\widehat{\left(\frac{\beta \cdot S}{N}\right)} = 0,01712$. оцінку величини

$$\frac{1}{\hat{\gamma}} \widehat{\left(\frac{\beta \cdot S}{N}\right)} = \widehat{\left(\frac{R_0 \cdot S}{N}\right)} = 1,9633. \quad (17)$$

Model is: $v2=a*\exp(b*v16)$ (Covid 19_2(Ukraine_2020))						
Dep. Var. : I						
Level of confidence: 95.0% (alpha=0.050)						
	Estimate	Standard error	t-value df = 211	p-value	Lo. Conf Limit	Up. Conf Limit
a	7833.070	87.14163	89.8889	0.00	7661.290	8004.849
b	0.019	0.00006	317.0763	0.00	0.019	0.019
Dep. Var. : Re						
Level of confidence: 95.0% (alpha=0.050)						
	Estimate	Standard error	t-value df = 211	p-value	Lo. Conf Limit	Up. Conf Limit
a	4951.668	113.3786	43.6738	0.00	4728.168	5175.168
b	0.017	0.0001	137.3798	0.00	0.017	0.017
Dep. Var. : I - Re						
Level of confidence: 95.0% (alpha=0.050)						
	Estimate	Standard error	t-value df = 211	p-value	Lo. Conf Limit	Up. Conf Limit
a	3086.667	72.11616	42.8013	0.00	2944.507	3228.828
b	0.020	0.00012	166.9584	0.00	0.020	0.020

Рис. 2. Параметри експоненціальної регресії змінних I , Re та $(I - Re)$ на номер доби з початку спостереження, отримані за допомогою програми Statistica 10

Незалежна змінна — номер доби ($v16$). Залежні змінні: I – Виявлено захворілих; Re – Всього вибуло (ті, хто одужав або помер); $(I - Re)$ – чисельність хворих.

Джерело: дані автора

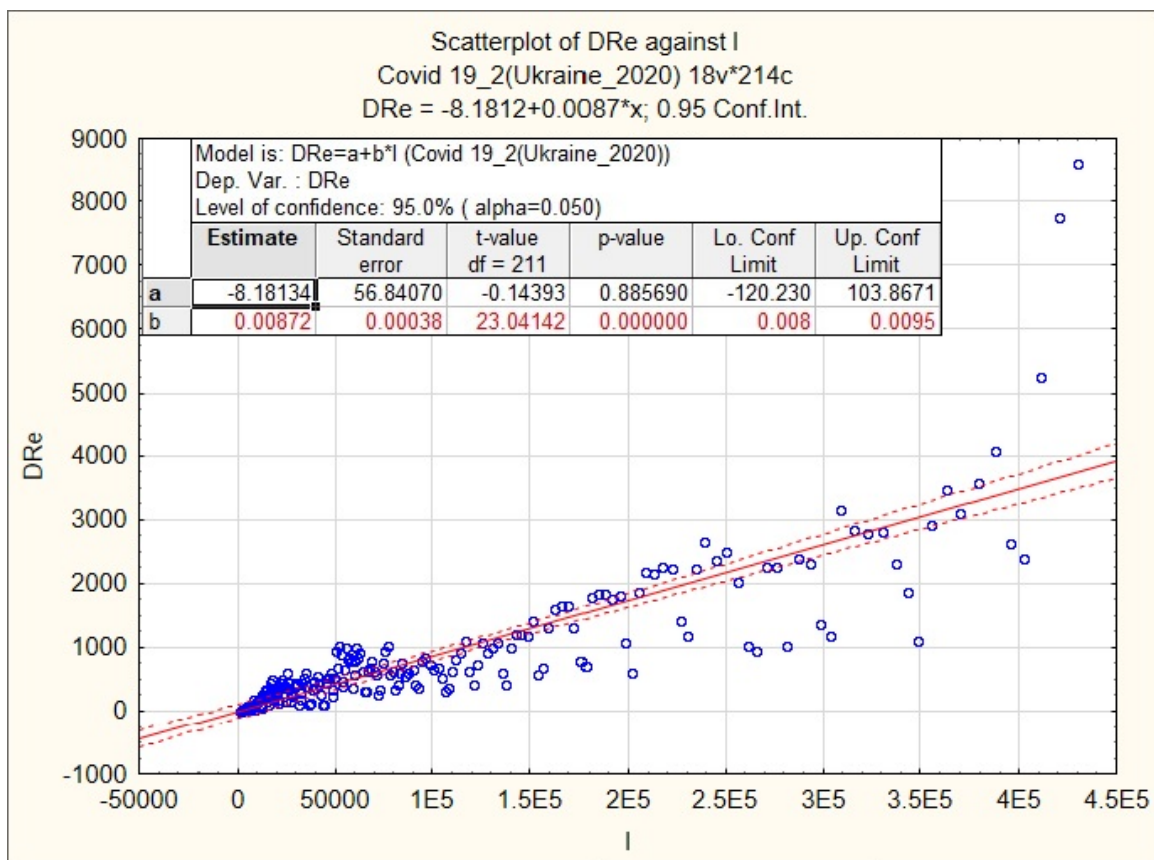


Рис. 4. Залежність чисельності тих, хто одужав і помер, від чисельності захворілих та її лінійна апроксимація

Джерело: дані автора

А тепер – головне. Якщо R_0 є сталою величиною, то частка незахворілих згідно з (17) також має бути однаковою в будь-який момент часу. Якщо виконується умова (12), це означає, що зі збільшенням чисельності інфікованих має зменшуватися кількість осіб, які вийшли з пандемії (одужали або померли). Поки що цього не спостерігається, тому приходимо до висновку, що *SIR*-модель не адекватна поточній динаміці пандемії. Чи працює *SEIR*-модель – задача наступного аналізу.

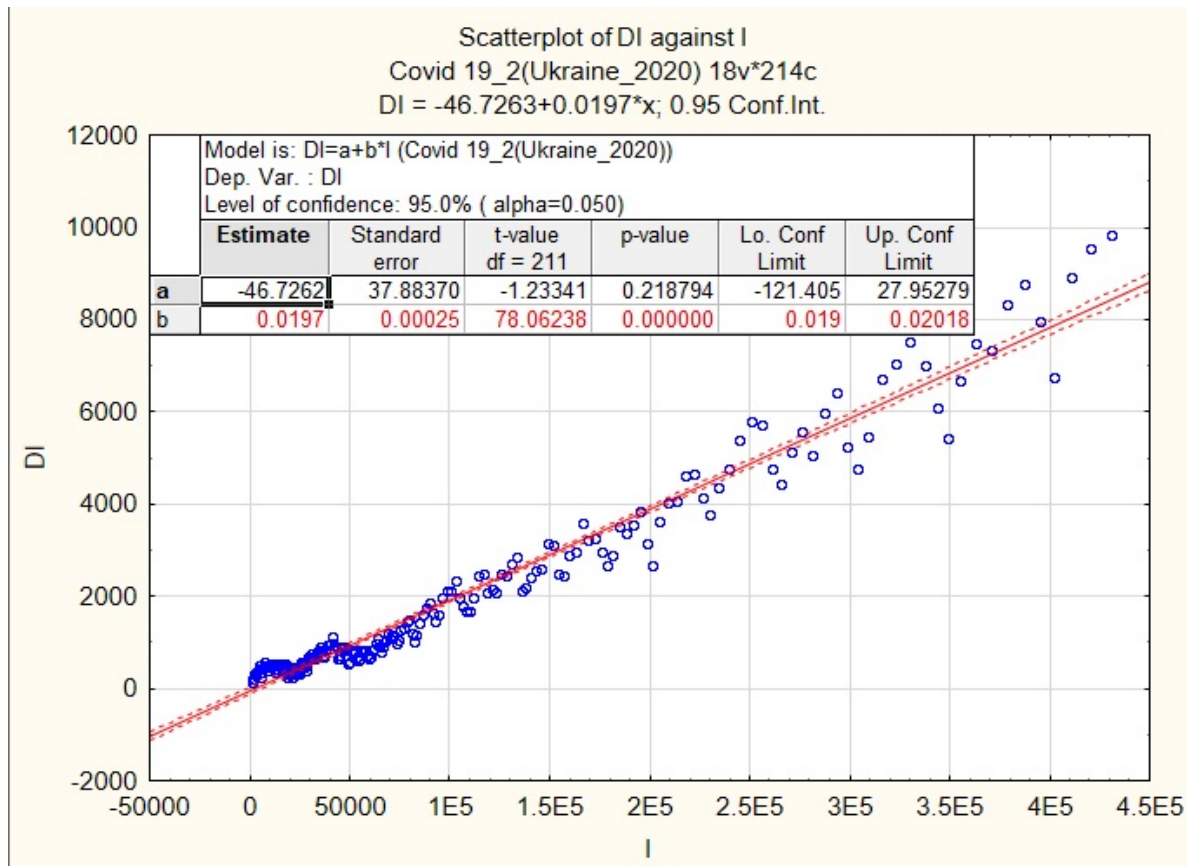


Рис. 5. Залежність добового збільшення чисельності захворілих від чисельності захворілих та її лінійна апроксимація

Джерело: дані автора

Подані результати отримано в рамках НДР «Дослідження деяких аспектів функціонування соціально-економічних систем в цифровій економіці» реєстраційний номер 0118U006677, 01.01.2019–31.12.2022.

Список використаних джерел

1. Бровченко, І. Розробка математичної моделі поширення епідемії COVID-19 в Україні. *Світгляд*. 2020. № 2(82). С. 2–14.
2. Kermack W. O., McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics. *Proc. Roy. Soc. Lond., Ser. A*. 1927. No 115. P. 700–721.
3. Вікіпедія. URL: <https://uk.wikipedia.org/wiki/> (дата звернення: 20.11.2020).
4. Зараза, гостя наша. Как математика помогает бороться с эпидемиями. URL: <https://nplus1.ru/material/2019/12/26/epidemic-math> (дата звернення: 20.11.2020).