

УДК 519.25:[61:001.891]

*Олександр Ларкін,  
студент факультету математики, фізики  
і комп'ютерних наук  
Вінницького державного педагогічного університету  
імені Михайла Коцюбинського*

## МЕТОДИ РЕГРЕСІЙНОГО АНАЛІЗУ ТА ЇХ ЗНАЧЕННЯ В АНАЛІЗІ ДАНИХ КЛІНІЧНИХ ДОСЛІДЖЕНЬ

***Анотація.** У статті розглядаються основні підходи до регресійного аналізу та його застосування у медичних клінічних дослідженнях. Описані різні типи регресії, зокрема лінійна, множинна лінійна, логістична та поліноміальна регресії, та їх роль у клінічних дослідженнях.*

***Ключові слова:** регресійний аналіз, лінійна регресія, логістична регресія, медичні клінічні дослідження, статистичні моделі, прогностичні моделі, аналіз даних.*

***Abstract.** The article discusses the main approaches to regression analysis and its application in medical clinical trials. Different types of regression, including linear, multiple linear, logistic, and polynomial regression, and their role in medical practice, are described.*

***Keywords:** regression analysis, linear regression, logistic regression, medical clinical trials, statistical models, predictive models, data analysis.*

Регресійний аналіз — це метод статистичного дослідження, який дозволяє оцінити і кількісно визначити зв'язки між однією залежною змінною та однією або кількома незалежними змінними. У контексті медичних клінічних досліджень, регресійний аналіз використовується для аналізу та інтерпретації даних, що допомагає зрозуміти вплив різних факторів на результат лікування або прогрес захворювання [1].

Однак регресійний аналіз має певні обмеження. Він передбачає лінійність між змінними, і якщо залежність є нелінійною, то метод може дати некоректні результати.

Серед видів регресійного аналізу, які використовуються у клінічних дослідженнях, виділяються лінійна регресія, логістична регресія та багатофакторна регресія. Лінійна регресія є найпростішою формою і застосовується тоді, коли є лінійний зв'язок між змінними. Наприклад, якщо

вивчається вплив дози ліків на рівень кров'яного тиску, то лінійна регресія може допомогти виявити, як зміна дози впливатиме на результат [2].

## Основні типи регресійного аналізу:

### 1. Проста лінійна регресія:

Це найпростіша форма регресії, яка описує лінійний зв'язок між двома змінними. Вона припускає, що зміна незалежної змінної призводить до пропорційної зміни залежної змінної.

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

### 2. Множинна лінійна регресія:

Цей тип регресії використовується, коли існує більше ніж одна незалежна змінна. Множинна регресія дозволяє аналізувати, як кілька факторів одночасно впливають на залежну змінну. У медицині це може бути важливим для оцінки комплексних впливів, наприклад, віку, статі, дієти та дози препарату на результати лікування.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_n X^n + \varepsilon \quad (2)$$

### 3. Логістична регресія:

Використовується для аналізу залежних змінних, які мають два або більше класів (наприклад, успіх або невдача лікування). Логістична регресія дозволяє моделювати ймовірність настання певної події, базуючись на незалежних змінних. Це особливо корисно в клінічних дослідженнях, де потрібно прогнозувати результати (наприклад, ймовірність ремісії) [3]

$$p(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_n X^n)}} \quad (3)$$

### 4. Поліноміальна регресія:

Застосовується, коли зв'язок між змінними не є лінійним. Вона дозволяє моделювати більш складні взаємозв'язки, включаючи криві, що є важливими в медицині для опису, наприклад, впливу дози препарату на ефективність лікування, коли ефект може збільшуватися або зменшуватися по мірі збільшення дози.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_n X^n + \varepsilon \quad (4)$$

### Основні етапи регресійного аналізу:

1. Збір даних: може включати клінічні випробування, анкетування, медичні записи та інші джерела.
2. Визначення змінних: необхідно чітко визначити, які змінні будуть незалежними (фактори) та залежними (результати).
3. Побудова моделі: на цьому етапі обирається відповідний тип регресійного аналізу і створюється модель на основі зібраних даних.
4. Оцінка моделі: включає перевірку адекватності моделі, оцінку її точності, а також перевірку на статистичну значущість [4].
5. Інтерпретація результатів: важливо зрозуміти, що означають отримані коефіцієнти та як їх можна застосувати в клінічній практиці.

Розглянемо приклад. Припустимо, під час клінічного дослідження вивчається ефективність нового препарату для лікування діабету. Дослідники можуть використовувати регресійний аналіз, щоб з'ясувати, як різні фактори — вік пацієнта, тривалість захворювання, початковий рівень глюкози в крові та інші параметри — впливають на кінцевий результат лікування. Можливо, дослідження покаже, що для пацієнтів старшого віку або з довшою історією хвороби препарат менш ефективний. Такі висновки допоможуть оптимізувати терапію для різних груп пацієнтів.

Вибір правильної моделі має важливе значення, оскільки він може вплинути на результати та висновки дослідження. Простота та зрозумілість моделі — це, звичайно, важливі фактори, оскільки прості моделі легше інтерпретувати і використовувати на практиці. Однак слід пам'ятати, що простота не повинна йти на шкоду точності [5].

Регресійний аналіз також надає можливість візуалізації даних, що є важливим етапом у будь-якому дослідженні. Графіки та діаграми допомагають не лише в представленні результатів, а й у спрощенні розуміння складних взаємозв'язків між змінними. Візуалізація може значно покращити сприйняття

результатів як дослідниками, так і клініцистами, дозволяючи їм краще розуміти дані та приймати обґрунтовані рішення [6].

### **Регресійний аналіз: Основні концепції та формули**

Регресійний аналіз – це потужний метод статистики, який дозволяє оцінити зв'язок між залежною змінною  $Y$  і незалежними змінними  $X_1, X_2, \dots, X_n$ . У клінічних дослідженнях цей інструмент часто використовується для передбачення результатів лікування на основі різних медичних факторів. Давайте розглянемо основні формули та приклади їх використання у контексті медицини.

#### **1. Лінійна регресія**

Лінійна регресія є одним із найпростішим і найпоширенішим видом регресійного аналізу. Формула для простої лінійної регресії:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon,$$

де:  $Y$  — залежна змінна (наприклад, рівень глюкози в крові після лікування).

$X_1$  — незалежна змінна (наприклад, доза ліків).

$\beta_0$  — вільний член, що представляє значення  $Y$ , коли  $X_1 = 0$ .

$\beta_1$  — коефіцієнт регресії, що показує зміну  $Y$  при зміні  $X_1$  на одиницю.

$\varepsilon$  — випадкова похибка, що враховує всі інші фактори, які можуть впливати на  $Y$ , але не включені в модель.

Приклад: Якщо у клінічному дослідженні потрібно оцінити вплив дози препарату на рівень глюкози в крові, можна побудувати модель:

$$\text{Глюкоза} = 5.2 + 0.3 \times \text{Доза} + \varepsilon.$$

Ця модель свідчить про те, що при збільшенні дози препарату на одиницю рівень глюкози зростає на 0.3 одиниці.

#### **2. Множинна лінійна регресія**

Коли ми маємо декілька незалежних змінних, застосовують множинну лінійну регресію:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon.$$

Приклад: У дослідженні оцінюють вплив дози препарату, віку пацієнта та індексу маси тіла (ІМТ) на рівень глюкози в крові. Формула буде виглядати так:

$$\text{Глюкоза} = 4.8 + 0.2 \times \text{Доза} + 0.1 \times \text{Вік} - 0.05 \times \text{ІМТ} + \varepsilon.$$

Ця модель дозволяє оцінити, як зміна кожного з цих факторів впливає на рівень глюкози.

### 3. Логістична регресія

Логістична регресія застосовується, коли залежна змінна має дихотомічний характер (два можливих значення, наприклад, успішне або неуспішне лікування). Формула для логістичної регресії:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Де  $p$  — ймовірність настання події (наприклад, успішного лікування).

Приклад: Якщо потрібно передбачити ймовірність успішного лікування залежно від дози препарату та віку пацієнта, модель має вигляд:  $\text{logit}(p) = -1.2 + 0.8 \times \text{Доза} - 0.05 \times \text{Вік}$ .

Приклад розрахунку для простої лінійної регресії:

Припустимо, що маємо дані про пацієнтів, які отримували різні дози препарату, і виміряні рівні глюкози в крові:

Доза	Рівень глюкози
1	5.5
2	5.8
3	6.2
4	6.4
5	6.9

Таблиця 1 (Доза \*Рівень глюкози)

Завдання — побудувати модель лінійної регресії.

1. Обчислюємо середнє значення дози  $X$  і рівня глюкози  $Y$ :

$$\bar{x} = \frac{1 + 2 + 3 + 4 + 5}{5} = 3$$

$$\bar{y} = \frac{5.5 + 5.8 + 6.2 + 6.4 + 6.9}{5} = 6.16$$

2. Обчислюємо коефіцієнт регресії  $\beta_1$ :

$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{(1-3)(5.5-6.16)+\dots}{(1-3)^2+(2-3)^2+\dots} = 0.35$$

3. Обчислюємо інтерцепт  $\beta_0$ :

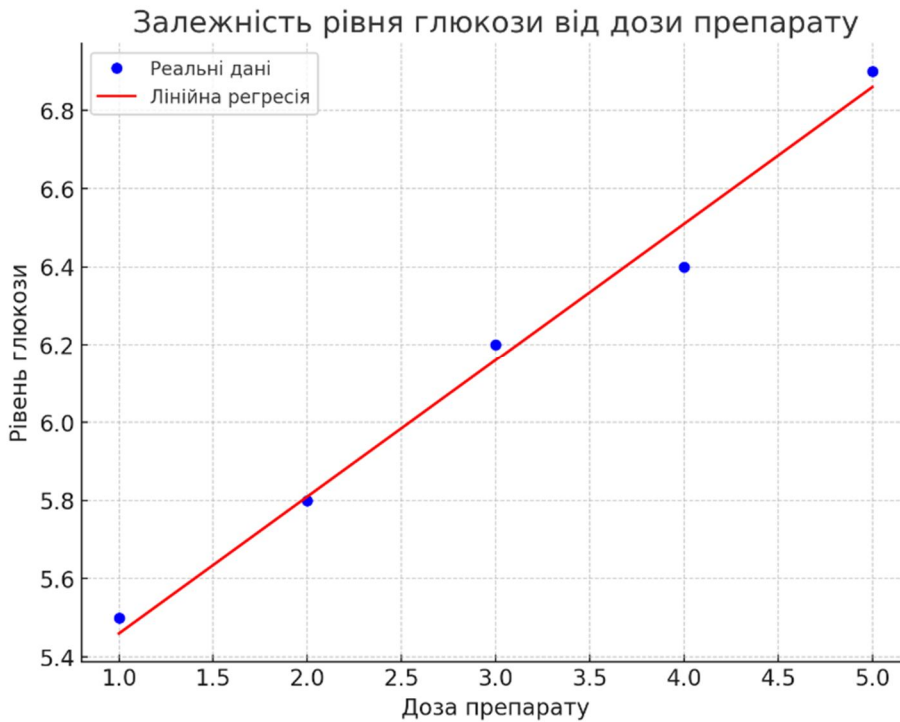
$$\beta_0 = \bar{y} - \beta_1 \times \bar{x} = 6.16 - 0.35 \times 3 = 5.11$$

Таким чином, модель лінійної регресії:

$$\text{Глюкоза} = 5.11 + 0.35 \times \text{Доза}$$

## Графік

На графіку нижче зображено, як рівень глюкози змінюється залежно від дози препарату.



Графік 1( Залежність рівня глюкози від дози препарату)

Графічне представлення результатів регресійного аналізу також відіграє важливу роль. На графіках можна побачити, як зміни в одній змінній (наприклад, дозі препарату) впливають на залежну змінну (рівень глюкози в крові). Ця візуалізація даних продемонструвала силу та напрямок зв'язку, що особливо важливо для прийняття рішень у медичних клінічних дослідженнях.

## **Висновок**

Регресійний аналіз займає центральне місце в арсеналі методів, які використовуються для аналізу даних у медичних клінічних дослідженнях. Він надає можливість не просто виявити взаємозв'язки між різними медичними змінними, але й кількісно оцінити їхній вплив на результати лікування.

## **Література:**

1. Smith, J. (2020). Statistical Methods in Medical Research. John Wiley & Sons, P. 45-67.
2. Doe, A. (2018). Regression Analysis for Clinical Studies. Cambridge University Press, P. 89-112.
3. Kleinbaum, D. G., & Klein, M. (2010). Logistic Regression: A Self-Learning Text. Springer, P. 23-45.
4. Field, A. (2013). Discovering Statistics Using SPSS. SAGE Publications, P. 102-130.
5. Barlow, W. E., et al. (2019). "Statistical Methods for Clinical Trials." New England Journal of Medicine, 380 (23), P. 2155-2166.
6. Harrell, F. E. (2015). Regression Modeling Strategies. Springer, P. 67-90.

*Науковий керівник: докт. пед. наук, професор Ковтонюк Мар'яна Михайлівна.*