КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (31), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.31.1023
УДК 004.056.5:519.23

Глушак Оксана Михайлівна
кандидат педагогічних наук, доцент,
доцент кафедри комп'ютерних наук
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID: 0000-0001-9849-1140
o.hlushak@kubg.edu.ua

Семеняка Світлана Олексіївна
кандидат фізико-математичних наук, доцент,
завідувач кафедри математики і фізики
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID: 0000-0001-5083-1433
s.semeniaka@kubg.edu.ua

Зінченко Надія Мусіївна
доктор фізико-математичних наук, старший науковий співробітник,
професор кафедри математики і фізики
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID: 0000-0003-1124-4636
nm.zinchenko@kubg.edu.ua

Соломко Вікторія Олександрівна
кандидат фізико-математичних наук,
доцент кафедри математики і фізики
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID: 0009-0004-0622-4977
v.solomko@kubg.edu.ua

ОЦІНЮВАННЯ ФАКТОРІВ СТІЙКОСТІ СИСТЕМ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ МЕТОДАМИ РЕГРЕСІЙНОГО АНАЛІЗУ

Анотація. У статті досліджуються методи регресійного аналізу як ефективний інструмент оцінювання факторів, що визначають стійкість систем інформаційної безпеки. Зокрема, показано, що застосування моделей множинної регресії дозволяє кількісно визначати вплив різноманітних технічних, організаційних та поведінкових чинників на рівень захищеності цифрових систем, прогнозувати ймовірність виникнення кіберзагроз, оцінювати ефективність заходів протидії та оптимізувати розподіл ресурсів захисту. Особлива увага приділена застосуванню методу найменших квадратів (МНК) для оцінювання параметрів регресійних моделей, а також вимогам, що забезпечують коректність отриманих результатів, зокрема незалежності факторних змінних, відсутності мультиколінеарності, гомоскедастичності та некорельованості залишків. У статті детально розглянуто діагностику основних статистичних порушень: мультиколінеарності, гетероскедастичності та автокореляції залишків та методи їх усунення, включаючи трансформацію змінних, введення допоміжних факторів або використання альтернативних підходів оцінювання, таких як узагальнений метод найменших квадратів (УМНК), метод головних компонент (РСА). Показано, що систематичне виявлення і корекція цих порушень є критично важливими для підвищення точності прогнозів і надійності висновків у сфері кібербезпеки, оскільки помилкові оцінки можуть призвести до недооцінки ризиків або неправильного визначення пріоритетів при управлінні загрозами. Робота демонструє, що статистично обґрунтоване застосування моделей регресійного аналізу створює міцну аналітичну основу для побудови інтелектуальних систем підтримки рішень у сфері інформаційної безпеки, підвищує ефективність моніторингу та сприяє формуванню стратегічних підходів до управління кіберризиками. Отримані результати можуть бути використані як у наукових дослідженнях, що спрямовані на вдосконалення методології оцінювання кіберризиків, так і при

практичному формуванні систем моніторингу та захисту інформаційних ресурсів, забезпечуючи підвищення рівня стійкості цифрової інфраструктури та ефективності заходів кіберзахисту.

Ключові слова: інформаційна безпека; регресійний аналіз; множинна регресія; мультиколінеарність; гетероскедастичність; автокореляція.

ВСТУП

Постановка проблеми. Стрімке зростання масштабів цифровізації, ускладнення інформаційних систем та еволюція кіберзагроз створюють нові виклики для забезпечення стійкості систем інформаційної безпеки. Зростання кількості інцидентів, багатофакторність атак та їхня непередбачуваність вимагають переходу від описових або експертних підходів до методів, що дозволяють здійснювати об'єктивний, кількісний аналіз впливу різних чинників на рівень захищеності цифрової інфраструктури. У таких умовах математичне моделювання, зокрема методи регресійного аналізу, набувають особливої значущості, оскільки забезпечують інструментарій для формалізації взаємозв'язків між характеристиками інформаційних систем, параметрами загроз та результативністю механізмів захисту.

Попри високий потенціал регресійних моделей для оцінювання стійкості систем інформаційної безпеки, їх застосування пов'язане з низкою методологічних складнощів. Множинна регресія потребує дотримання ключових статистичних передумов, зокрема відсутності мультиколінеарності, автокореляції та гетероскедастичності, порушення яких призводить до спотворення оцінок і зниження надійності прогнозів. У контексті кібербезпеки це може мати суттєві наслідки, оскільки неточність параметрів моделі здатна спричинити неправильне визначення пріоритетів захисту, недооцінку ризиків або неефективний розподіл ресурсів. Тому постає проблема обґрунтованого використання методів регресійного аналізу для оцінювання факторів стійкості систем інформаційної безпеки, що включає не лише побудову моделі, а й коректну діагностику та усунення статистичних порушень або застосування альтернативних методів оцінювання.

Аналіз останніх досліджень і публікацій. Статистичні методи математичного моделювання посідають центральне місце серед аналітичних підходів, оскільки дають змогу кількісно описувати залежності між реальними показниками та оцінювати силу впливу різних факторів. Одним із ключових інструментів такого аналізу є регресійне моделювання, яке забезпечує побудову емпірично обґрунтованих залежностей між змінними. У випадках, коли досліджуване явище формують кілька взаємопов'язаних чинників, найбільш інформативними стають моделі множинної регресії – вони дозволяють комплексно виявити структуру впливів, визначити провідні фактори та підвищити точність прогнозів [1-3].

У науковій практиці такі моделі активно застосовуються в економічних, фінансових, природничих і технічних дослідженнях. Вони слугують основою аналізу ризиків, прогнозування складних процесів, оптимізації ресурсів, а також підтримують прийняття управлінських рішень у багатофакторних системах [4]. У різних галузях (від медицини до логістики) множинна регресія довела свою універсальність та високу аналітичну цінність.

В умовах цифрової трансформації значення математичного моделювання суттєво зростає [5-7]. Перехід суспільних і економічних процесів у цифрове середовище супроводжується збільшенням обсягів даних і появою нових загроз: кібератак, порушень

цілісності інформації, витоків конфіденційних даних. Забезпечення кібербезпеки сьогодні потребує не лише технічних засобів, а й глибокого розуміння факторів, що визначають уразливість інформаційних систем [8-11]. Саме тому регресійні моделі стають важливим інструментом аналізу закономірностей виникнення та розвитку кіберзагроз.

Методи множинної регресії дозволяють оцінювати вплив технічних, організаційних і поведінкових чинників на рівень стійкості систем інформаційної безпеки, визначати ймовірність настання інцидентів, ідентифікувати приховані залежності між характеристиками мережевого середовища. Їх використання сприяє формуванню доказової основи для стратегій кіберзахисту, оптимізації систем моніторингу та підвищення адаптивності цифрової інфраструктури.

З огляду на це, наукові праці останніх років дедалі більше зосереджуються на поєднанні регресійного аналізу з технологіями машинного навчання, інтелектуального аналізу даних та автоматизованих систем підтримки рішень у сфері безпеки. Огляд цих досліджень є ключовим для розуміння потенціалу регресійних моделей у прогнозуванні кібератак і підвищенні ефективності заходів захисту. Зокрема, з'являється тенденція до інтеграції класичних статистичних підходів із сучасними алгоритмами глибинного навчання та ансамблевого моделювання, що дає змогу підвищувати точність оцінювання ризиків і адаптивність аналітичних систем до нових типів загроз. На цьому тлі особливу увагу привертають дослідження, у яких регресійні моделі розглядаються як складова гібридних інтелектуальних систем безпеки, здатних одночасно забезпечувати інтерпретованість результатів та високу прогностичну здатність [12].

Попри те, що публікації [13-17] зосереджені на глибинному навчанні, підкріпленому навчанні та гібридних інтелектуальних моделях, їхні підходи концептуально спираються на логіку регресійного аналізу – насамперед у частині моделювання залежностей, оцінювання впливу предикторів і побудови прогнозних функцій. Використання часових рядів у LSTM-мережах, ймовірнісних оцінок безпеки через softmax, а також геометричного поділу рішень у просторі ознак фактично є нелінійними узагальненнями класичних регресійних підходів, що дозволяють враховувати складні взаємозв'язки між змінними.

У сучасних дослідженнях, присвячених машинному навчанню в інженерії та кібербезпеці, окрему увагу приділено вдосконаленню методів прогнозування на основі високовимірних даних, де регресійні моделі виступають ключовою складовою аналітичної інфраструктури. Навіть у складних гібридних або автономних системах (від структурного аналізу до виявлення кіберзагроз) саме регресійні підходи забезпечують формалізацію залежностей, оцінювання ризиків і побудову прогнозних функцій, на які накладаються більш складні алгоритми. Таким чином, сучасні інтелектуальні рішення не витісняють регресійні моделі, а навпаки розширюють їх застосування, використовуючи їх як базовий інструмент для аналізу, калібрування параметрів та підвищення точності предиктивних систем [17-18].

У дослідженнях, присвячених інтелектуальним системам підтримки рішень, навіть за використання складних архітектур, фундаментальне значення зберігають регресійні моделі, які забезпечують базове оцінювання залежностей, трендів і часових закономірностей у даних. Саме регресійний підхід дозволяє формалізувати вплив ключових факторів, задати напрями оптимізації стратегії та створити інтерпретовану основу, на якій будуються більш складні алгоритми глибокого підкріплювального навчання. [19-20].

Таким чином, застосування моделей множинної регресії у сфері інформаційної та кібернетичної безпеки є не лише науково обґрунтованим, але й суспільно необхідним. Воно сприяє формуванню аналітичних основ для розроблення політик захисту даних, вдосконалення систем моніторингу кіберзагроз та підвищення рівня стійкості цифрової інфраструктури. У поєднанні з сучасними інформаційними технологіями математичне моделювання стає потужним інструментом забезпечення безпеки інформаційного простору, що визначає стратегічні напрями розвитку як національних, так і міжнародних систем кіберзахисту.

Мета статті. Метою статті є обґрунтування можливостей застосування методів регресійного аналізу, насамперед моделей множинної регресії, для кількісного оцінювання факторів, що визначають стійкість систем інформаційної та кібернетичної безпеки. У межах дослідження передбачається окреслити математичні засади використання методу найменших квадратів, визначити умови його коректного застосування та проаналізувати ефективні підходи до виявлення й усунення статистичних порушень (мультиколінеарності, гетероскедастичності та автокореляції) з метою підвищення достовірності та практичної цінності отриманих моделей.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Метод найменших квадратів (МНК) є одним із базових статистичних підходів до оцінювання параметрів регресійних моделей. Його головна ідея полягає у мінімізації суми квадратів відхилень фактичних значень результативної змінної від теоретичних, отриманих на основі моделі. За дотримання певних передумов МНК забезпечує отримання незміщених, ефективних і найменш дисперсних оцінок коефіцієнтів регресії. Саме ці властивості роблять метод найменших квадратів ключовим інструментом у побудові аналітичних залежностей для складних багатофакторних процесів, зокрема, у сфері кібербезпеки, де необхідно досліджувати вплив багатьох змінних (технічних, організаційних, поведінкових) на рівень захищеності інформаційних систем.

Застосування МНК є коректним лише за умови виконання низки статистичних вимог, які визначають адекватність і надійність побудованої моделі. Основні з них такі:

1. *Випадковість залишків і нульове математичне сподівання похибок.* Кожне значення складової u_i регресійного рівняння $y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n + u$ має розглядатися як випадкова величина, а математичне сподівання залишків повинно дорівнювати нулю: $M(u) = 0$.

Це означає, що модель не має систематичних зсувів у прогнозуванні, тобто, вона не «завищує» і не «занижує» результати. Для задач кібербезпеки така умова є важливою, оскільки дозволяє уникнути упереджених оцінок ризиків або ймовірностей інцидентів, що могло б призвести до неправильного розподілу ресурсів захисту.

2. *Некорельованість залишків і сталість дисперсії (гомоскедастичність).* Компоненти вектора залишків повинні бути статистично незалежними і мати однакову дисперсію: $Cov(u_i, u_j) = 0$, $i \neq j$; $D(u) = M(u_i - Mu_i)^2 = \sigma_{u_i}^2 = const$, $i = \overline{1, n}$.

Це означає, що випадкові відхилення не мають системної структури, а рівень похибок не змінюється залежно від обсягу чи природи даних. У дослідженнях кібербезпеки це гарантує, що результати аналізу не залежать від масштабу мережевих систем або частоти виявлення атак і не спотворюють загальну картину ризику.

3. *Некорельованість факторних змінних із залишками.* Змінні X_i, які використовуються як предиктори, не повинні бути пов'язані з похибками моделі: $Cov(X_i, u_i) = 0$.

Ця вимога забезпечує об'єктивність оцінок: пояснювальні фактори не мають містити інформації, що вже відображена у випадковій складовій. У задачах оцінювання рівня інформаційної безпеки це дозволяє уникнути ситуацій, коли неконтрольовані фактори (наприклад, людський чинник або непередбачувані зовнішні впливи) спотворюють результати моделювання.

4. *Відсутність мультиколінеарності між факторними змінними.* Пояснювальні змінні не повинні бути тісно лінійно пов'язаними між собою, тобто: $Cov(X_i, X_j) \approx 0,\ i \neq j$.

Висока кореляція між факторами призводить до нестабільності оцінок коефіцієнтів і ускладнює їх інтерпретацію. У сфері кібербезпеки це особливо важливо, адже різні показники (наприклад, частота спроб несанкціонованого доступу, кількість оновлень програмного забезпечення, рівень завантаження мережі) можуть бути взаємопов'язаними. Правильне виявлення й усунення мультиколінеарності дозволяє підвищити точність прогнозів і достовірність висновків.

Дотримання наведених передумов забезпечує надійність статистичних висновків і дає змогу інтерпретувати отримані коефіцієнти моделі як реальні кількісні оцінки впливу окремих факторів на досліджуваний результат. У контексті аналізу систем інформаційної та кібернетичної безпеки це означає можливість створення формалізованих моделей ризику, прогнозування вразливостей або визначення ключових чинників, що впливають на ефективність заходів захисту.

На практиці ідеальні умови застосування МНК виконуються рідко. Реальні дані, особливо в інформаційно-технологічних і кібернетичних системах, часто містять випадкові або систематичні спотворення, що призводять до відхилення від теоретичних передумов. Тому одним із найважливіших етапів побудови регресійної моделі є виявлення та усунення порушень базових статистичних припущень, які можуть суттєво вплинути на достовірність оцінок параметрів. Серед таких порушень найбільш поширеними є мультиколінеарність, гетероскедастичність та автокореляція.

Мультиколінеарність означає наявність сильного лінійного зв'язку між двома або більше пояснювальними змінними. У цьому випадку модель «втрачає» здатність чітко розподіляти вплив кожного фактора, що призводить до великих стандартних похибок оцінок коефіцієнтів і нестабільності результатів при найменших змінах у вихідних даних. Для систем кібербезпеки, де фактори часто взаємопов'язані (наприклад, кількість зареєстрованих атак, навантаження на сервери, обсяг переданих даних тощо), ігнорування мультиколінеарності може призвести до хибних висновків про значущість тих чи інших показників. Тому на етапі побудови моделі необхідно проводити діагностику мультиколінеарності, наприклад, за допомогою показників інфляції дисперсії або коефіцієнтів парної кореляції, і в разі потреби – виключати або об'єднувати надлишкові змінні.

У практичних дослідженнях важливо не лише встановити факт наявності мультиколінеарності, а й обрати відповідний інструментарій для її виявлення. Існує низка статистичних методів та критеріїв, що дають змогу оцінити ступінь взаємозалежності пояснювальних змінних та визначити, наскільки серйозною є проблема для конкретної моделі. Основні з них узагальнено у таблиці «Методи діагностики мультиколінеарності та їх характеристика».

Таблиця 1

Методи діагностики мультиколінеарності та їх характеристика

Метод діагностики	Переваги / сильні сторони	Недоліки / обмеження
Коефіцієнти парної кореляції між змінними	Простота обчислення та інтерпретації. Дозволяє швидко виявити очевидні лінійні зв'язки між парами змінних.	Не виявляє мультиколінеарність, що виникає внаслідок взаємодії багатьох змінних. Чутливий лише до парних лінійних зв'язків.
Матриця кореляцій	Наглядне графічне представлення структури зв'язків. Зручна для попереднього аналізу багатьох змінних одночасно.	Як і парна кореляція, не виявляє приховану мультиколінеарність вищого порядку.
Метод головних компонент (PCA)	Дозволяє зменшити розмірність даних і виявити групи тісно пов'язаних змінних. Дає можливість сформувати нові незалежні компоненти.	Складніший в інтерпретації. Заміна вихідних змінних компонентами може ускладнити трактування результатів регресії.
Індекс інфляції дисперсії (VIF)	Найпоширеніший і найбільш інформативний критерій. Виявляє мультиколінеарність як між окремими змінними, так і у взаємодії багатьох факторів. Добре інтерпретується.	Вимагає попереднього оцінювання моделі. Може давати завищені значення при наявності нелінійних зв'язків або аномалій у даних.
Детермінант кореляційної матриці	Дає узагальнену оцінку рівня мультиколінеарності всієї системи змінних. Корисний при великій кількості факторів.	Не дозволяє визначити, які саме змінні створюють проблему. Чутливий до масштабування змінних.
Критерій Фаррара–Глобера	Статистично обґрунтований комплексний тест. Виявляє глобальну, часткову та парну мультиколінеарність.	Складніший у застосуванні й потребує більшого обсягу розрахунків. Менш інтуїтивний для початкового аналізу.

Гетероскедастичність виникає тоді, коли дисперсія залишків змінюється залежно від значень пояснювальних змінних, що свідчить про нерівномірність варіації похибок у різних частинах вибірки. За таких умов оцінки коефіцієнтів, отримані методом найменших квадратів, хоч і залишаються незміщеними, проте втрачають ефективність: стандартні похибки стають неточними, а довірчі інтервали – нерепрезентативними. У сфері інформаційної безпеки подібна проблема може проявлятися під час різких коливань мережевого трафіку або пікових навантажень, коли моделі помилково «згладжують» реальні ризики кібератак.

Для виявлення неоднорідності дисперсії використовуються спеціалізовані статистичні критерії, які дають змогу оцінити її наявність і ступінь прояву. Найпоширеніші методи діагностики узагальнено в таблиці «Методи діагностики гетероскедастичності та їх характеристика».

Таблиця 2

Методи діагностики гетероскедастичності та їх характеристика

Метод діагностики	Переваги / сильні сторони	Недоліки / обмеження
Графічний аналіз	Простий, інтуїтивно зрозумілий підхід. Дозволяє візуально виявити різні форми гетероскедастичності (конусоподібність, "віяло", зміни розкиду).	Не дає кількісної оцінки. Може бути неточним при великому шумі або значних відхиленнях.

Тест Голдфельда–Квандта	Один із класичних тестів; добре працює при підозрі на монотонну зміну дисперсії. Простий у застосуванні та інтерпретації.	Вимагає розбиття вибірки на групи, що може спотворити результати. Малоефективний при складних формах гетероскедастичності.
Тест Бройша–Пагана	Дозволяє кількісно оцінити залежність дисперсії від однієї або кількох змінних. Підходить для різних моделей і великих вибірок.	При нелінійній залежності від факторів може бути нечутливим. Чутливий до порушень нормальності залишків.
Тест Уайта	Універсальний тест без припущень про форму гетероскедастичності. Дозволяє виявляти як лінійні, так і нелінійні залежності дисперсії.	Може сигналізувати про гетероскедастичність там, де її немає (підвищена чутливість). Потребує порівняно великих вибірок для стабільних оцінок.
Модифікований тест Уайта	Дозволяє коригувати стандартні помилки без усунення самої гетероскедастичності. Підвищує надійність статистичних висновків.	Не усуває гетероскедастичність, лише мінімізує її вплив. Менш ефективний при дуже малих вибірках.
Аналіз абсолютних або квадратичних залишків	Простий у виконанні та дозволяє оцінити характер залежності дисперсії від факторів. Корисний при припущенні про певну форму гетероскедастичності.	Малоефективний, якщо форма залежності невідома. Чутливий до вибросів та порушень нормальності.

Автокореляція має місце тоді, коли залишки регресійної моделі виявляють систематичну залежність між сусідніми спостереженнями. Найчастіше це характерно для часових рядів, де динаміка процесів формується послідовно в часі. За таких умов стандартні похибки, отримані методом найменших квадратів, стають некоректними, що спотворює результати статистичних тестів і може призвести до хибних висновків про значущість окремих факторів. У дослідженнях, пов’язаних з інформаційною безпекою, подібна ситуація є поширеною, адже інтенсивність кіберзагроз, частота мережевих атак чи рівень навантаження на інфраструктуру часто змінюються поступово, утворюючи корельовану структуру похибок.

Для виявлення такого типу порушення застосовують спеціальні статистичні тести, що дозволяють оцінити характер і силу залежності між залишками. Найвідоміші методи діагностики узагальнено в таблиці «Методи діагностики автокореляції та їх характеристика».

Таблиця 3

Методи діагностики автокореляції та їх характеристика

Метод діагностики	Переваги / сильні сторони	Недоліки / обмеження
Критерій Дарбіна-Уотсона (DW-тест)	Найбільш відомий і стандартний тест для перевірки першого порядку автокореляції. Простий в обчисленні. та інтерпретації. Підходить для класичних часових рядів.	Перевіряє лише автокореляцію першого порядку. Не застосовується при наявності лагової змінної залежної змінної у моделі
Тест Бройша-Годфрі	Більш універсальний, ніж DW-тест. Дозволяє тестувати автокореляцію вищих порядків. Можна застосовувати в моделях з лагами залежної змінної.	Чутливий до специфікаційних помилок моделі. Потребує правильно обраного порядку автокореляції.
Q-тест Льюнга-Бокса	Перевіряє автокореляцію одночасно для кількох лагів. Широко використовується в аналізі часових рядів. Добре працює з великими вибірками.	Чутливий до вибору кількості лагів. Менш ефективний на малих вибірках.

Тест фон Неймана	Простий та використовується для базової оцінки наявності автокореляції між послідовними залишками	Менш чутливий, ніж DW-тест. Не дозволяє досліджувати автокореляцію вищих порядків
Аналіз спектральної щільності залишків	Дозволяє виявити циклічні компоненти автокореляції. Корисний у складних часових моделях (кібератаки, мережевий трафік).	Складний у застосуванні. Потребує високої якості даних та специфічних інструментів для аналізу.
Кореляція між залишками та лагами регресорів (Durbin's *h*-test)	Ефективний тоді, коли DW-тест не можна застосувати. Дає змогу оцінити вплив лагів залежної змінної на структуру помилок.	Вузькоспеціалізований тест. Чутливий до мультиколінеарності між лагами.

Таким чином, виявлення та корекція мультиколінеарності, гетероскедастичності й автокореляції є ключовою умовою побудови статистично обґрунтованої та практично придатної регресійної моделі. У сфері інформаційної та кібернетичної безпеки, де дані характеризуються високою динамічністю, нерівномірністю та наявністю прихованих залежностей, ігнорування цих порушень може призвести до хибної оцінки рівня загроз або неправильної інтерпретації впливу критичних факторів. Тому важливо не лише своєчасно діагностувати статистичні відхилення, але й застосовувати методи, здатні мінімізувати їхній вплив.

У ситуаціях, коли неоднорідність дисперсії не усувається стандартними засобами (перетвореннями змінних чи зміною функціональної форми), ефективним є використання зваженого методу найменших квадратів (ЗМНК), що дозволяє адаптувати модель до нерівномірності варіації залишків. Аналогічно, при наявності стійкої автокореляції, підтвердженої критерієм Дарбіна-Уотсона або тестом Бройша-Годфрі, доцільно застосовувати методи узагальнених найменших квадратів (УМНК) чи моделі, що враховують часову структуру даних, зокрема авторегресивні підходи. Такі інструменти підвищують точність оцінювання параметрів і забезпечують коректність аналітичних висновків, що є критично важливим для прийняття обґрунтованих рішень щодо захисту інформаційних ресурсів та управління кіберризиками.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Проведений аналіз підтвердив, що регресійні методи, зокрема моделі множинної регресії, є потужним інструментом для кількісного вивчення факторів, які визначають стійкість систем інформаційної безпеки. На відміну від дескриптивних або виключно експертних підходів, ці моделі дають змогу об'єктивно оцінювати силу впливу окремих показників, виявляти приховані залежності в даних та формувати прогностичні оцінки, важливі для виявлення потенційних кіберзагроз і планування заходів захисту.

Ефективність використання методу найменших квадратів у регресійному аналізі залежить від дотримання фундаментальних статистичних вимог, що забезпечують надійність отриманих оцінок. Порушення таких передумов, зокрема появи корельованих залишків, неоднорідності дисперсії та надмірної взаємозалежності пояснювальних змінних, призводить до втрати точності та стабільності моделі. У дослідженнях кібербезпеки, де аналітичні результати є основою стратегічних і тактичних рішень, така втрата коректності може мати критичні наслідки для оцінювання ризиків та ефективності протидії атакам.

Діагностика статистичних відхилень є обов'язковою складовою моделювання. Такі інструменти, як VIF, тести Бройша-Пагана і Уайта, критерій Дарбіна-Уотсона та тест

Бройша-Годфрі, дозволяють своєчасно виявляти структурні проблеми у даних і обирати відповідні методи корекції. Якщо ж повністю усунути їх не вдається, застосування альтернативних підходів, наприклад, узагальненого МНК або методу головних компонент (PCA), забезпечує збереження стійкості оцінювання і підвищує адекватність моделей у складних умовах кіберсередовища.

Забезпечення статистичної цілісності моделей має ключове значення для формування достовірних прогнозів і побудови систем підтримки рішень у сфері безпеки інформаційних ресурсів. Коректно специфіковані моделі сприяють точнішому виявленню закономірностей у поведінці загроз, мінімізують ризик хибних інтерпретацій та підвищують ефективність аналітичних платформ моніторингу й раннього реагування на інциденти. Математично виважене моделювання створює підґрунтя для впровадження інтелектуальних технологій підтримки рішень у сфері управління інформаційними ризиками та кіберзахисту

Перспективи подальших досліджень полягають у розширенні спектра математичних і статистичних методів, придатних для аналізу складних та динамічних даних кіберсередовища. Зокрема, доцільним є поєднання класичних моделей множинної регресії з методами машинного навчання, байєсівськими підходами та нелінійними моделями для виявлення прихованих залежностей у великих обсягах даних. Подальшого розвитку потребують також алгоритми адаптивного оцінювання параметрів у реальному часі, що дозволить підвищити чутливість моделей до змін у кіберпросторі. Важливим напрямом подальших досліджень є інтеграція регресійних методів у системи моніторингу безпеки та автоматизовані рішення з оцінювання ризиків, що забезпечить глибше розуміння процесів, пов’язаних із загрозами, уразливостями та стійкістю цифрової інфраструктури.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Abramov, V., Astafieva, M., Boiko, M., Bodnenko, D., Bushma, A., Vember, V., Hlushak, O., Zhyltsov, O., Ilich, L., Kobets, N., Kovaliuk, T., Kuchakovska, H., Lytvyn, O., Lytvyn, P., Mashkina, I., Morze, N., Nosenko, T., Proshkin, V., Radchenko, S., … Yaskevych, V. (2021). *Theoretical and practical aspects of the use of mathematical methods and information technology in education and science*. https://doi.org/10.28925/9720213284km.
2. Vasechko L., Hlushak, O., Semenyaka, S., Ramskyi, A., & Nesterova, O. (2022). Diagnosis profitable part of the pension fund of Ukraine by method of mathematical modeling. *Financial and Credit Activity Problems of Theory and Practice*, 2(43), 157-164. https://doi.org/10.55643/fcaptp.2.43.2022.3638
3. Ilich, L., Hlushak, O., Semenyaka, S.(2020) Modeling of employment structural transformations. *Financial and credit activity: problems of theory and practice*. ISSN (print) 2306-4994, ISSN (on-line) 2310-8770, 1(32), P.251-259.
4. Ilich L., Hlushak O., Semenyaka, S. & Shestack, Y. (2023) «Modeling of Structural Changes in the Employment as the Direction of Economic Security Risk Management» *Cybersecurity Providing in Information and Telecommunication Systems*, (3421.) p p. 46-55. ISSN 1613-0073.
5. Карпунін, І., Зінченко, Н. (2023). Когнитивне моделювання інтелектуальних систем аналізу фінансового стану суб’єкта господарювання *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*, 1(21), 75–85.
6. Astafieva, M., Stepanenko, N. (2025) Optimal Cybersecurity Management: Global Controllability of Linear Models *Cybersecurity Providing in Information and Telecommunication Systems* (3991). c. 14-25. ISSN 1613-0073
7. Astafieva, M., Stepanenko, N. (2025) Оптимальне управління кібербезпекою: глобальна керованість лінійних моделей *Cybersecurity Providing in Information and Telecommunication Systems* (3991). c. 14-25. ISSN 1613-0073)

8. Shevchenko, S., Zhdanova, Y., Spasiteleva, S., Mazur , N., Skladannyi, P., & Nehodenko, V. (2024). Mathematical methods in cyber security: cluster analysis and its application in information and cybernetic security. *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*, 3(23), 258-273. https://doi.org/10.28925/2663-4023.2024.23.258273
9. Shevchenko, S., Zhdanova, Y., & Spasiteleva, S. (2023). Mathematical methods in cybersecurity: Catastrophe theory. Cybersecurity: Education, Science, Technique, 3(19), 165–175. https://doi.org/10.28925/2663-4023.2023.19.165175
10. Shevchenko, S., Zhdanova Y., Skladannyi, P., & Spasiteleva, S. (2021). Mathematical methods in cibernetic security: graphs and their application in information and cybernetic security. *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*, 1(13), 133–144. https://doi.org/10.28925/2663-4023.2021.13.133144
11. Shevchenko, S., Zhdanova Y., Spasiteleva, S., Negodenko, O., Mazur, N., & Kravchuk, K. (2019). Mathematical methods in cyber security: fractals and their applications in information and cyber security. *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*, *1*(5), 31–39. https://doi.org/10.28925/2663-4023.2019.5.3139
12. Sitnikova, O., Melnyk, M., Syrota, O., & Semenov, S. (2025). Intelligent method for supporting decision-making on software security using hybrid models. *Innovative technologies and scientific solutions for industries* 1(31),115-126. https://doi.org/10.30837/2522-9818.2025.1.115.
13. Samia, N., Saha, S., & Haque, A. (2024). Predicting and mitigating cyber threats through data mining and machine learning. *Computer Communications*. Volume 228, 107949 https://doi.org/10.1016/j.comcom.2024.107949.
14. Salman, A., Al-Nuaimi, B., Subhi, A., Alkattan, H., & Alfilh, R. (2025). Enhancing Cybersecurity with Machine Learning: A Hybrid Approach for Anomaly Detection and Threat Prediction. *Mesopotamian Journal of CyberSecurity*. *5*(1), 202-215. https://doi.org/10.58496/mjcs/2025/014.
15. Gongada, T., Agnihotri, A., Santosh, K., Ponnuswamy, V., S, N., Sharma, T., & El-Ebiary, Y. (2024). Leveraging Machine Learning for Enhanced Cyber Attack Detection and Defence in Big Data Management and Process Mining. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/ijacsa.2024.0150266.
16. Alanazi, F., Alshammari, A., Sallami, C., Alhashmi, A., Effghi, R., Km, A., & Darem, A. (2025). Enhancing cybersecurity vulnerability detection using different machine learning severity prediction models. *International Journal of Applied Science and Engineering*. 22, 2025013. https://doi.org/10.6703/ijase.202503_22(1).003.
17. Sarker, I. (2022). Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. *Annals of Data Science*, 10, 1473 - 1498. https://doi.org/10.1007/s40745-022-00444-2.
18. Kumar, S., Machireddy, J., Sankaran, T., & Sholapurapu, P. (2025). Integration of Machine Learning and Data Science for Optimized Decision-Making in Computer Applications and Engineering. *Journal of Information Systems Engineering and Management*. 10 №45s, https://doi.org/10.52783/jisem.v10i45s.8990.
19. Li, P., & Zhang, L. (2025). Application of big data technology in enterprise information security management. *Scientific Reports*, 15, 1022. https://doi.org/10.1038/s41598-025-85403-6.
20. Lu, M. (2025). Automated Machine Learning and Data-Driven Decision Support System for Strategy Management in Organizational Activities. *J. Comput. Inf. Technol.*, 1(33), 57-71. https://doi.org/10.20532/cit.2025.1005877.

.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (31), 2025
ISSN 2663 - 4023

Oksana Hlushak
PhD, Associate Professor,
Associate Professor of the Department of Computer Science
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0000-0001-9849-1140
o.hlushak@kubg.edu.ua

Svitlana Semeniaka
PhD, Associate Professor,
Head of the Department of Mathematics and Physics
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0000-0001-5083-1433
s.semeniaka@kubg.edu.ua

Nadiia Zinchenko
Doctor of Science, senior research fellow,
Professor of the Department of Mathematics and Physics
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0000-0003-1124-4636
nm.zinchenko@kubg.edu.ua

Viktoriia Solomko
PhD, Associate Professor of the Department of Mathematics and Physics
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0009-0004-0622-4977
v.solomko@kubg.edu.ua

ASSESSMENT OF STABILITY FACTORS OF INFORMATION SECURITY SYSTEMS USING REGRESSION ANALYSIS METHODS

Abstract. The article explores regression analysis methods as an effective tool for assessing factors that determine the stability of information security systems. In particular, it is shown that the use of multiple regression models allows us to quantitatively determine the impact of various technical, organizational and behavioral factors on the level of security of digital systems, predict the probability of cyber threats, assess the effectiveness of countermeasures and optimize the allocation of protection resources. Particular attention is paid to the application of the least squares method (LSM) to estimate the parameters of regression models, as well as the requirements that ensure the correctness of the obtained results, in particular, the independence of factor variables, the absence of multicollinearity, homoscedasticity, and uncorrelatedness of residuals. The article examines in detail the diagnostics of the main statistical violations: multicollinearity, heteroscedasticity and autocorrelation of residuals and methods for their elimination, including the transformation of variables, the introduction of auxiliary factors or the use of alternative estimation approaches, such as the generalized least squares method (GLS), the principal component analysis (PCA). It is shown that the systematic detection and correction of these violations are critically important for increasing the accuracy of forecasts and the reliability of conclusions in the field of cybersecurity, since erroneous assessments can lead to underestimation of risks or incorrect prioritization in threat management. The work demonstrates that the statistically sound application of regression analysis models creates a solid analytical basis for building intelligent decision support systems in the field of information security, increases the effectiveness of monitoring and contributes to the formation of strategic approaches to cyber risk management. The results obtained can be used both in scientific research aimed at improving the methodology for assessing cyber risks, and in the practical formation of monitoring and protection systems for information resources, ensuring an increase in the level of resilience of digital infrastructure and the effectiveness of cyber protection measures.

Keywords: information security; regression analysis; multiple regression; multicollinearity; heteroscedasticity; autocorrelation.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Abramov, V., Astafieva, M., Boiko, M., Bodnenko, D., Bushma, A., Vember, V., Hlushak, O., Zhyltsov, O., Ilich, L., Kobets, N., Kovaliuk, T., Kuchakovska, H., Lytvyn, O., Lytvyn, P., Mashkina, I., Morze, N., Nosenko, T., Proshkin, V., Radchenko, S., … Yaskevych, V. (2021). *Theoretical and practical aspects of the use of mathematical methods and information technology in education and science*. https://doi.org/10.28925/9720213284km.
2. Vasechko L., Hlushak, O., Semenyaka, S., Ramskyi, A., & Nesterova, O. (2022). Diagnosis profitable part of the pension fund of Ukraine by method of mathematical modeling. *Financial and Credit Activity Problems of Theory and Practice*, 2(43), 157-164. https://doi.org/10.55643/fcaptp.2.43.2022.3638
3. Ilich, L., Hlushak, O., Semenyaka, S.(2020) Modeling of employment structural transformations. *Financial and credit activity: problems of theory and practice*. ISSN (print) 2306-4994, ISSN (on-line) 2310-8770, 1(32), P.251-259.
4. Ilich L., Hlushak O., Semenyaka, S. & Shestack, Y. (2023) «Modeling of Structural Changes in the Employment as the Direction of Economic Security Risk Management» *Cybersecurity Providing in Information and Telecommunication Systems*, (3421.) p p. 46-55. ISSN 1613-0073.
5. Karpunin, I., Zinchenko, N. (2023). Cognitive modeling of intelligent systems for analyzing the financial condition of a business entity *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*, 1(21), 75–85.
6. Astafieva, M., Stepanenko, N. (2025) Optimal Cybersecurity Management: Global Controllability of Linear Models *Cybersecurity Providing in Information and Telecommunication Systems* (3991). c. 14-25. ISSN 1613-0073
7. Astafieva, M., Stepanenko, N. (2025) Оптимальне управління кібербезпекою: глобальна керованість лінійних моделей *Cybersecurity Providing in Information and Telecommunication Systems* (3991). c. 14-25. ISSN 1613-0073)
8. Shevchenko, S., Zhdanova, Y., Spasiteleva, S., Mazur , N., Skladannyi, P., & Nehodenko, V. (2024). Mathematical methods in cyber security: cluster analysis and its application in information and cybernetic security. *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*, 3(23), 258-273. https://doi.org/10.28925/2663-4023.2024.23.258273
9. Shevchenko, S., Zhdanova, Y., & Spasiteleva, S. (2023). Mathematical methods in cybersecurity: Catastrophe theory. Cybersecurity: Education, Science, Technique, 3(19), 165–175. https://doi.org/10.28925/2663-4023.2023.19.165175
10. Shevchenko, S., Zhdanova Y., Skladannyi, P., & Spasiteleva, S. (2021). Mathematical methods in cibernetic security: graphs and their application in information and cybernetic security. *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*, 1(13), 133–144. https://doi.org/10.28925/2663-4023.2021.13.133144
11. Shevchenko, S., Zhdanova Y., Spasiteleva, S., Negodenko, O., Mazur, N., & Kravchuk, K. (2019). Mathematical methods in cyber security: fractals and their applications in information and cyber security. *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*, *1*(5), 31–39. https://doi.org/10.28925/2663-4023.2019.5.3139
12. Sitnikova, O., Melnyk, M., Syrota, O., & Semenov, S. (2025). Intelligent method for supporting decision-making on software security using hybrid models. *Innovative technologies and scientific solutions for industries*,1(31),115-126. https://doi.org/10.30837/2522-9818.2025.1.115.
21. Samia, N., Saha, S., & Haque, A. (2024). Predicting and mitigating cyber threats through data mining and machine learning. *Computer Communications*. Volume 228, 107949 https://doi.org/10.1016/j.comcom.2024.107949.
13. Salman, A., Al-Nuaimi, B., Subhi, A., Alkattan, H., & Alfilh, R. (2025). Enhancing Cybersecurity with Machine Learning: A Hybrid Approach for Anomaly Detection and Threat Prediction. *Mesopotamian Journal of CyberSecurity*. 5(1), 202-215. https://doi.org/10.58496/mjcs/2025/014.
14. Gongada, T., Agnihotri, A., Santosh, K., Ponnuswamy, V., S, N., Sharma, T., & El-Ebiary, Y. (2024). Leveraging Machine Learning for Enhanced Cyber Attack Detection and Defence in Big Data Management and Process Mining. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/ijacsa.2024.0150266.
15. Alanazi, F., Alshammari, A., Sallami, C., Alhashmi, A., Effghi, R., Km, A., & Darem, A. (2025). Enhancing cybersecurity vulnerability detection using different machine learning severity prediction

models. *International Journal of Applied Science and Engineering*. 22, 2025013. https://doi.org/10.6703/ijase.202503_22(1).003.

16. Sarker, I. (2022). Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. *Annals of Data Science*, 10, 1473 - 1498. https://doi.org/10.1007/s40745-022-00444-2.
17. Kumar, S., Machireddy, J., Sankaran, T., & Sholapurapu, P. (2025). Integration of Machine Learning and Data Science for Optimized Decision-Making in Computer Applications and Engineering. *Journal of Information Systems Engineering and Management*., 10 №45s, https://doi.org/10.52783/jisem.v10i45s.8990.
18. Li, P., & Zhang, L. (2025). Application of big data technology in enterprise information security management. *Scientific Reports*, 15, 1022. https://doi.org/10.1038/s41598-025-85403-6.
19. Lu, M. (2025). Automated Machine Learning and Data-Driven Decision Support System for Strategy Management in Organizational Activities. *J. Comput. Inf. Technol.*, 1(33), 57-71. https://doi.org/10.20532/cit.2025.1005877.