

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

«На правах рукопису»
УДК 004:519.24:681.3.06

«До захисту допущено»

Завідувач кафедри

_____ Едуард ЖАРІКОВ

«__» _____ 2024 р.

Магістерська дисертація

на здобуття ступеня магістра

**за освітньо-професійною програмою «Інженерія програмного забезпечення
інформаційних систем»**

зі спеціальності 121 «Інженерія програмного забезпечення»

**на тему: «Математичне та програмне забезпечення для побудови
багатовимірної лінійної регресії на малому об'ємі експериментальних
даних»**

Виконав (-ла):

студент (-ка) II курсу, групи ПІ-32мп

Грицюк Володимир Васильович _____

Керівник:

проф. кафедри ПІ, д.т.н., проф.

Павлов Олександр Анатолійович _____

Рецензент:

доцент кафедри ІСТ, к.т.н.

Солдатова Марія Олександрівна _____

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів без
відповідних посилань.

Студент (-ка) _____

Київ – 2024 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

Рівень вищої освіти – другий (магістерський)

Спеціальність – 121 «Інженерія програмного забезпечення»

Освітньо-професійна програма «Інженерія програмного забезпечення інформаційних систем»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Едуард ЖАРІКОВ

«__» _____ 2024р.

ЗАВДАННЯ
на магістерську дисертацію студенту
Грицюку Володимирі Васильовичу

- 1) Тема дисертації «Математичне і програмне забезпечення для побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних», науковий керівник дисертації Павлов Олександр Анатолійович, д.т.н., проф., проф. кафедри ІІІ, затверджені наказом по університету від «08» листопада 2024 р. № 5016-с.
- 2) Термін подання студентом дисертації «16» грудня 2024 р.
- 3) Об'єкт дослідження – математичне та програмне забезпечення пошуку коефіцієнтів багатовимірної лінійної регресії, заданої надлишковим описом.
- 4) Предмет дослідження – методи і програмні засоби побудови багатовимірної лінійної регресії та методологія дослідження їх ефективності на малому об'ємі експериментальних даних.
- 5) Перелік завдань, які потрібно розробити – критичний науковий аналіз методів побудови багатовимірних регресій; критичний аналіз програмного забезпечення, що реалізує методи побудови багатовимірних регресій; розробка методології проведення статистичного імітаційного експерименту моделювання ефективності оригінального методу побудови багатовимірної лінійної регресії, заданої надлишковим описом, по невеликом об'єму експериментальних даних; розробка та обґрунтування архітектури програмного забезпечення, що реалізує оригінальний метод побудови багатовимірної лінійної регресії, заданої

надлишковим описом, та системи статистичного імітаційного моделювання, що дозволяє знайти область його ефективного застосування; розробка програмного забезпечення, що реалізує оригінальний метод побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних та систему статистичного імітаційного моделювання його ефективності; розробка рекомендацій по використанню досліджуваного методу на основі проведених статистичних досліджень.

б) Орієнтовний перелік публікацій – одна публікація

7) Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

8) Дата видачі завдання «1» вересня 2024 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання	Примітка
1	Аналіз методів побудови БЛР	1.09.2024	
2	Аналіз ПЗ, що реалізує методи побудови БЛР	14.09.2024	
3	Розробка методології дослідження ефективності оригінального методу	01.10.2024	
4	Розробка бібліотеки, що реалізує оригінальний метод	10.10.2024	
5	Розробка системи статистичного імітаційного моделювання ефективності досліджуваного методу	20.10.2024	
6	Тестування ПЗ	31.10.2024	
7	Виконання експериментальних досліджень	10.11.2024	
8	Оформлення пояснювальної записки	20.11.2024	
9	Подання дисертації на попередній захист	22.11.2024	
10	Подання дисертації на захист	9.12.2024	

Студент

Володимир ГРИЦЮК

Науковий керівник

Олександр ПАВЛОВ

РЕФЕРАТ

Розмір пояснювальної записки – 126 аркушів, містить 20 ілюстрацій, 102 таблиці, 4 додатків, 30 посилань на джерела.

Актуальність теми. Актуальність проблеми побудови багатовимірних регресій на основі активного чи пасивного експерименту є актуальною, по перше, тому, що вона дозволяє знаходити детерміновані закономірності, що використовуються в сучасних інформаційних діагностичних системах, по друге, тому що в вихідні дані експерименту адитивно входять реалізації випадкової величини з достатньо великою дисперсією. Це приводить до того, що на сьогоднішній день ефективних універсальних методів побудови багатовимірних регресій не існує. Кожен з відомих методів є ефективним лише при певних обмеженнях, які задаються теоретично чи експериментально. Особлива складність побудови багатовимірних регресій зв'язана з об'ємом експериментальних даних, а саме чим менша їх кількість, тим складніше отримати ефективні оцінки коефіцієнтів багатовимірної регресії. Тому, створення метода побудови багатовимірної лінійної регресії заданої надлишковим описом по невеликому об'єму експериментальних даних, що підвищує ефективність загальної алгоритмічної процедури методу найменших квадратів є актуальною.

Мета дослідження. Основною метою є підвищення ефективності побудови багатовимірної лінійної регресії заданої надлишковим описом на малому об'ємі експериментальних даних.

Об'єкт дослідження: математичне та програмне забезпечення пошуку коефіцієнтів багатовимірної лінійної регресії заданої надлишковим описом.

Предмет дослідження: методи і програмні засоби побудови багатовимірної лінійної регресії та методологія дослідження їх ефективності на малому об'ємі експериментальних даних.

Для реалізації поставленої мети **сформульовані наступні завдання:**

– критичний науковий аналіз методів побудови багатовимірних регресій;

- критичний аналіз програмного забезпечення, що реалізують методи побудови багатовимірних регресій;
- розробка методології проведення статистичного імітаційного експерименту моделювання ефективності методу побудови багатовимірної лінійної регресії, заданої надлишковим описом, по невеликому об'єму експериментальних даних;
- розробка та обґрунтування архітектури програмного забезпечення, що реалізує оригінальний метод побудови багатовимірних лінійних регресій, заданих надлишковим описом, та системи статистичного імітаційного моделювання, що дозволяє знайти область його ефективного застосування;
- розробка програмного забезпечення, що реалізує оригінальний метод побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних та систему статистичного імітаційного моделювання його ефективності;
- розробка рекомендацій по використанню досліджуваного методу на основі проведених статистичних досліджень.

Наукова новизна результатів магістерської дисертації полягає в тому, що:

- розроблена оригінальна методологія проведення статистичних імітаційних експериментів для обґрунтування області ефективного використання універсального методу побудови багатовимірних лінійних регресій, заданих надлишковим описом, по невеликому об'єму експериментальних даних;
- розроблена та обґрунтована структура програмного забезпечення, що вперше реалізує оригінальний метод побудови багатовимірної лінійної регресії, заданої надлишковим описом по малому об'єму експериментальних даних та систему статистичного імітаційного моделювання для знаходження області його ефективного використання.

Практичне значення отриманих результатів полягає в тому, що розроблено кросплатформенну бібліотеку (моноліт), що реалізує оригінальний алгоритм побудови багатовимірної лінійної регресії, заданої надлишковим

описом, на малому об'ємі експериментальних даних, яка може використовуватись у сучасних діагностичних і експертних системах.

Зв'язок з науковими програмами, планами, темами. Робота виконувалась на кафедрі інформатики та програмної інженерії Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського".

Апробація. Наукові положення дисертації пройшли апробацію на VII Міжнародній науково-практичній конференції молодих вчених та студентів «Інженерія програмного забезпечення і передові інформаційні технології SoftTech-2024».

Публікації. Наукові положення дисертації опубліковані в:

1) Павлов О. А., Головченко М. М., Грицюк В. В. Методологія дослідження ефективності метода побудови багатовимірної лінійної регресії по малому об'єму експериментальних даних // Інженерія програмного забезпечення і передові інформаційні технології (SoftTech-2024): матеріали тез доповідей VII Міжнародної науково-практичної конференції молодих вчених та студентів – м. Київ: НТУУ «КПІ ім. Ігоря Сікорського» 19-22 листопада 2024р.

Ключові слова: БАГАТОВИМІРНА ЛІНІЙНА РЕГРЕСІЯ, МЕТОД НАЙМЕНШИХ КВАДРАТІВ, ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ.

ABSTRACT

Explanatory note size – 126 pages, contains 20 illustrations, 102 tables, 4 applications, 30 references.

Topicality. The relevance of the problem of constructing multivariate regressions based on an active or passive experiment is actual, firstly, because it allows finding deterministic patterns used in modern information diagnostic systems, and secondly, because the initial data of the experiment additively include realizations of a random variable with a sufficiently large variance. This leads to the fact that today there are no effective universal methods for constructing multivariate regressions. Each of the known methods is effective only under certain restrictions that are set theoretically or experimentally. The particular complexity of constructing multivariate regressions is associated with the volume of experimental data, namely, the smaller their number, the more difficult it is to obtain effective estimates of the coefficients of multivariate regression. Therefore, the creation of a method for constructing multivariate linear regression given by an over-description on a small volume of experimental data, which increases the efficiency of the general algorithmic procedure of the least squares method, is relevant.

The aim of the study. The main target is to increase the efficiency of constructing multivariate linear regression given by over-description on a small amount of experimental data.

The object of research: mathematical and software solutions for calculating coefficients of multivariate linear regression given by over-description.

The subject of research: methods and software tools for constructing multivariate linear regression and methodology for studying their effectiveness on a small amount of experimental data.

To achieve this goal, the **following tasks** were formulated:

- critical scientific analysis of methods for constructing multivariate regressions;
- critical analysis of software implementing methods for constructing multivariate regressions;

- development of a methodology for conducting a statistical simulation experiment to model the effectiveness of the method for constructing multivariate linear regressions given by over-description on a small amount of experimental data;
- development and justification of the architecture of software that implements the original method for constructing multivariate linear regressions given by over-description, and a statistical simulation modeling system that allows finding the area of its effective application;
- development of software that implements the original method for constructing multivariate linear regressions on a small amount of experimental data and a system for statistical simulation modeling of its effectiveness;
- development of recommendations for using the method under study based on the conducted statistical studies.

The scientific novelty of the results of the master's dissertation is:

- an original methodology for conducting statistical simulation experiments was developed to substantiate the area of effective use of the universal method for constructing multivariate linear regressions, given by over-description, on a small amount of experimental data;
- a software structure has been developed and substantiated, which for the first time implements an original method for constructing multivariate linear regression given by an over-description on a small volume of experimental data and a statistical simulation modeling system for finding the area of its effective use.

The practical value of the obtained results is that a cross-platform library (monolith) has been developed that implements an original algorithm for constructing multivariate linear regression, given by an over-description, on a small amount of experimental data, which can be used in modern diagnostic and expert systems.

Relationship with working with scientific programs, plans, topics. Work was performed at the Department of Informatics and Software Engineering of the National Technical University of Ukraine «Kyiv Polytechnic Institute. Igor Sikorsky».

Approbation. The scientific provisions of the dissertation were tested at the VII International Scientific and Practical Conference of Young Scientists and Students “Software Engineering and Advanced Information Technologies SoftTech-2024”.

Publications. The scientific provisions of the dissertation were published in:

1) Pavlov O. A., Golovchenko M. M., Hrytsiuk V. V. Methodology for studying the effectiveness of the method for constructing multivariate linear regression on a small volume of experimental data // Software Engineering and Advanced Information Technologies (SoftTech-2024): materials of the abstracts of the VII International Scientific and Practical Conference of Young Scientists and Students – Kyiv: NTUU “KPI them Igor Sikorsky” November 19-22, 2024.

Keywords: MULTIDIMENSIONAL LINEAR REGRESSION, LEAST SQUARES METHOD, SIMULATION MODELING.

ЗМІСТ

ВСТУП	14
1 АНАЛІЗ МЕТОДІВ ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ПОБУДОВИ БАГАТОВИМІРНОЇ РЕГРЕСІЇ	16
1.1 Опис предметної області.....	16
1.2 Критичний аналіз існуючих методів побудови багатовимірних регресій	18
1.2.1 Метод максимальної правдоподібності	18
1.2.2 Множинний дискримінантний аналіз	19
1.2.3 Метод найменших квадратів	20
1.2.4 Факторний аналіз.....	21
1.2.5 Метод групового урахування аргументів	22
1.3 Критичний аналіз програмного забезпечення, що реалізує методи побудови багатовимірних регресій	23
1.3.1 Scikit-learn.....	23
1.3.2 NumPy	24
1.3.3 R-Stats	24
1.3.4 STATISTICA	25
1.3.5 SPSS Statistics	25
1.4 Постановка задачі	25
Висновки до розділу	27
2 ОРИГІНАЛЬНИЙ МЕТОД ПОБУДОВИ БАГАТОВИМІРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ НА МАЛОМУ ОБ'ЄМІ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ	28
2.1 Опис методу підвищення продуктивності загальної процедури МНК.....	28
2.1.1 Постановка задачі.....	28
2.1.2 Етапи методу.....	28
2.2 Методологія статистичного дослідження ефективності оригінального методу побудови БЛР, заданої надлишковим описом	29
2.2.1 Параметри моделювання, що змінюються в процесі статистичних досліджень:	30
2.2.2 Процедура статистичного імітаційного моделювання.....	30
Висновки до розділу	31

3 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ	32
3.1 Опис вимог до програмного забезпечення.....	32
3.2 Розробка архітектури програмного забезпечення	33
3.3 Опис програмного забезпечення	35
3.4 Графічний інтерфейс користувача	39
3.5 Інструкція з розгортання програмного забезпечення	47
Висновки до розділу	49
4 МОДЕЛЮВАННЯ ЕФЕКТИВНОСТІ ЗАСТОСУВАННЯ	
ДОСЛІДЖУВНОГО МЕТОДА ЗА ДОПОМОГОЮ РОЗРОБЛЕНОГО	
ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ.....	50
4.1 Процедура проведення досліджень.....	50
4.2 Результати статистичного моделювання ефективності застосування	
досліджуваного метода побудови БЛР при рівномірному розподілі випадкової	
величини E	51
4.3 Результати статистичного моделювання ефективності застосування	
досліджуваного метода побудови БЛР при нормальному розподілі випадкової	
величини E	77
4.4 Результати статистичного моделювання ефективності застосування	
досліджуваного метода побудови БЛР при сумі нормального і рівномірного	
розподілу величини E	93
Висновки до розділу	104
5 МАРКЕТИНГОВИЙ АНАЛІЗ СТАРТАП-ПРОЄКТУ.....	106
5.1 Опис ідеї проєкту	106
5.2 Технологічний аудит ідеї проєкту.....	107
5.3 Аналіз ринкових можливостей запуску стартап-проєкту.....	108
5.4 Розроблення ринкової стратегії проєкту	113
5.5 Розробка маркетингової програми стартап-проєкту.....	117
Висновки до розділу	119
ВИСНОВКИ	120
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	122
ДОДАТОК А.....	126
ДОДАТОК Б	127

ДОДАТОК В.....	128
ДОДАТОК Г	139

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БЛР – Багатовимірна лінійна регресія

МНК – Метод найменших квадратів

ІТ – інформаційні технології

ВСТУП

Актуальність проблеми побудови багатовимірних регресій на основі активного чи пасивного експерименту є актуальною, по перше, тому, що вона дозволяє знаходити детерміновані закономірності, що використовуються в сучасних інформаційних діагностичних системах, по друге, тому, що у вихідні дані експерименту адитивно входять реалізації випадкової величини з достатньо великою дисперсією. Це приводить до того, що на сьогоднішній день ефективних універсальних методів побудови багатовимірних регресій не існує. Кожен з відомих методів є ефективним лише при певних обмеженнях, які задаються теоретично чи експериментально. Особлива складність побудови багатовимірних регресій пов'язана з об'ємом експериментальних даних, а саме чим менша їх кількість, тим складніше отримати ефективні оцінки коефіцієнтів багатовимірної регресії. Тому, створення метода побудови багатовимірної лінійної регресії заданої надлишковим описом по невеликому об'єму експериментальних даних, що підвищує ефективність загальної алгоритмічної процедури методу найменших квадратів є актуальною.

Основною метою є підвищення ефективності побудови багатовимірної лінійної регресії заданої надлишковим описом на малому об'ємі експериментальних даних.

Об'єктом дослідження є математичне та програмне забезпечення пошуку коефіцієнтів багатовимірної регресії, заданої надлишковим описом.

Предметом дослідження є процес розробки програмного забезпечення, що реалізує оригінальний алгоритм побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних.

Для досягнення поставленої мети було визначено наступні завдання:

- критичний науковий аналіз методів побудови багатовимірних регресій;
- критичний аналіз програмного забезпечення, що реалізують методи побудови багатовимірних регресій;

- розробка методології проведення статистичного імітаційного експерименту дослідження ефективності методу побудови багатовимірної лінійної регресії, заданої надлишковим описом, по невеликому об'єму експериментальних даних;
- розробка та обґрунтування архітектури програмного забезпечення, що реалізує оригінальний метод побудови багатовимірних лінійних регресій, заданих надлишковим описом, та системи статистичного імітаційного моделювання, що дозволяє знайти область його ефективного застосування;
- розробка програмного забезпечення, що реалізує оригінальний метод побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних та систему статистичного імітаційного моделювання його ефективності;
- розробка рекомендацій по використанню досліджуваного методу на основі проведених статистичних досліджень.

Наукова новизна результатів магістерської дисертації полягає в тому, що:

- розроблена оригінальна методологія проведення статистичних імітаційних експериментів для обґрунтування області ефективного використання універсального методу побудови багатовимірних лінійних регресій, заданих надлишковим описом, по невеликому об'єму експериментальних даних;
- розроблена та обґрунтована структура програмного забезпечення, що вперше реалізує оригінальний метод побудови багатовимірної лінійної регресії, заданої надлишковим описом по малому об'єму експериментальних даних та систему статистичного імітаційного моделювання для знаходження області його ефективного використання.

Практичне значення отриманих результатів полягає в тому, що розроблено красплатформенну бібліотеку (моноліт), що реалізує оригінальний алгоритм побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних, яка може використовуватись у сучасних діагностичних і експертних системах.

1 АНАЛІЗ МЕТОДІВ ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ПОБУДОВИ БАГАТОВИМІРНОЇ РЕГРЕСІЇ

1.1 Опис предметної області

Регресійний аналіз – метод, що застосовується для ідентифікації взаємозв'язків між залежною та незалежними змінними [29]. Для регресійного аналізу не важливо, що саме відбувається всередині досліджуваної системи, він сприймає її як чорний ящик. На вхід такої системи подаються вектори незалежних змінних X , а в результаті, отримується набір параметрів Y [29]. Таким чином, математична модель залежності набору вихідних змінних Y та вхідних параметрів X виражається формулою:

$$Y = F(X) + E, \quad (1.1)$$

де X – набір незалежних, вхідних змінних, Y – набір залежних, вихідних змінних, E – випадкова похибка. Окрім того, вхідні та вихідні змінні мають враховувати наступні обмеження:

- вимірювання значень незалежних змінних X має проводитись з високою точністю, більшою, ніж точність вимірювання значень результуючих змінних Y ;
- змінні серед набору незалежних змінних X мають бути незалежні один від одного;
- рекомендується застосовувати активний експеримент – значення вхідних змінних X задаються зовні (якщо це можливо);
- набір незалежних змінних X повинен формуватись таким чином, щоб кожна змінна з цього набору впливає на значення результуючої змінної Y .

Всеосяжне визначення багатовимірної регресії [28, 29] – множина поліномів, що використовується для отримання багатовимірної регресії за допомогою лінійної зваженої згортки деякої підмножини цих поліномів.

Також математична модель регресії включає в себе деяку похибку, що відображає теоретично невраховані інші параметри системи.

Регресії поділяються на одновимірні та багатовимірні [29]. У свою чергу, одновимірні регресії поділяються на поліноміальні та гіперболічні [29]. Багатовимірні регресії поділяються на поліноміальні та логістичні [29]. Одновимірні регресії застосовуються тоді, коли значення вихідної змінної залежить тільки від значення однієї вхідної змінної. Багатовимірна регресія застосовується тоді, коли значення вихідної змінної залежить від значень декількох вхідних змінних.

Математична модель одновимірної поліноміальної регресії n -порядку представлено далі [7, 20]:

$$k_0 + k_1x + \dots + k_nx^n + E, \quad (1.2)$$

де x – незалежна змінна, $k_0 \dots k_n$ – коефіцієнти регресії, E – випадкова похибка.

Математична модель одновимірної гіперболічної регресії представлено далі [3]:

$$Y(x) = k_0 + k_1 \cdot \frac{1}{x} + E, \quad (1.3)$$

де x – незалежна змінна, k_0 та k_1 – коефіцієнти регресії, E – випадкова похибка.

Зазвичай, на практиці значно частіше досліджують залежність вихідної змінної від деякого набору вхідних змінних. В такому випадку застосовується багатовимірна регресія. Математична модель багатовимірної поліноміальної регресії наведена далі [4, 15, 16]:

$$Y(\underline{X}) = \sum_{i,j=1}^{i,j=k} k_{i_1 \dots i_m}^{j_1 \dots j_m} (x_{i_1})^{j_1} \cdot (x_{i_2})^{j_2} \cdot \dots \cdot (x_{i_m})^{j_m} + E$$

$$\forall (i_1 \dots i_t) \in K, \forall (j_1 \dots j_t) \in K(i_1 \dots i_t), \quad (1.4)$$

де X – набір, вектор вхідних, незалежних змінних, E – випадкова похибка.

Часто стоїть задача визначити, до якого класу відноситься спостережуване значення. Цю задачу допомагає вирішити логістична регресія. Математична модель багатовимірної логістичної регресії наведена далі [2]:

$$y^*(\underline{X}) = k_0 + k_1x_1 + k_2x_2 + \dots + k_mx_m + E,$$

$$Y = \{1, y^* \leq 0, y^* > 0\} \quad (1.5)$$

де X – набір, вектор вхідних, незалежних змінних, E – випадкова похибка.

У даній роботі буде досліджуватися багатовимірна поліноміальна регресія.

1.2 Критичний аналіз існуючих методів побудови багатовимірних регресій

Розглянемо універсальні методи, що можуть використовуватися для побудови регресійних моделей, серед яких метод максимальної правдоподібності, множинний дискримінантний аналіз, метод найменших квадратів, факторний аналіз, метод групового урахування аргументів.

1.2.1 Метод максимальної правдоподібності

Метод максимальної правдоподібності – це метод, що застосовується для пошуку параметрів регресійної моделі за допомогою максимізації імовірності спостережень на конкретному наборі вхідних даних. У математичній статистиці, даний метод є фундаментом для пошуку коефіцієнтів регресійної моделі базуючись на вибірці даних [11, 12, 13]. Для побудови регресійної моделі, метод максимальної правдоподібності допомагає у пошуку таких коефіцієнтів, щоб відображення вхідних змінних у вихідні було найточніше.

Застосування методу максимальної подібності розпочинається з пошуку функції правдоподібності, що описує ймовірність значень вихідних змінних за тих чи інших значень вхідних змінних. Часто для спрощення подальших обчислень, отримана функція логарифмується. Таким чином, операції множення підмінюються операціями додавання, а операції ділення – операціями віднімання. Отримана функція максимізується – проводиться пошук параметрів, при яких значення функції правдоподібності найбільші. Таким чином, знайдені параметри функції максимальної правдоподібності і будуть коефіцієнтами регресійної моделі. Приклад застосування методу максимальної подібності для задачі одновимірної поліноміальної регресії показано на рисунку 1.1.

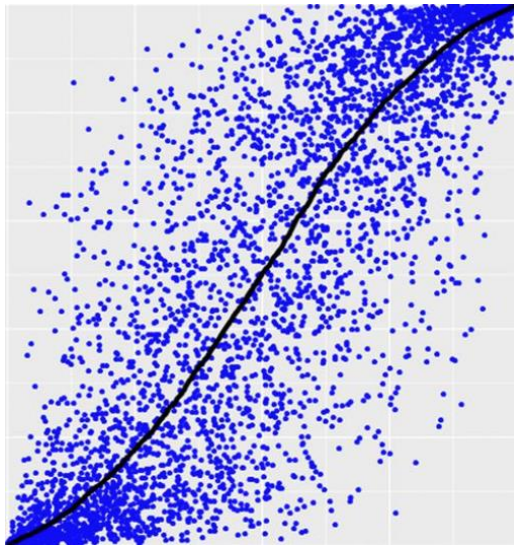


Рисунок 1.1 – Приклад застосування методу максимальної подібності для задачі одновимірної поліноміальної регресії [26]

1.2.2 Множинний дискримінантний аналіз

Множинний дискримінантний аналіз – це метод побудови регресійних моделей, в основі якого лежить розділ даних на кілька підгруп ґрунтуючись на значеннях конкретних змінних чи наборів змінних з метою мінімізації варіації всередині підгрупи та максимізації варіації між підгрупами [6, 9, 21].

Для розподілу даних за підгрупами алгоритм множинного дискримінантного аналізу використовує дискримінантну функцію – таку комбінацію змінних, щоб можна було як найкраще розподілити дані на підгрупи (класи). Виходячи з такої специфіки, в регресійному аналізі множинний дискримінантний аналіз використовується для побудови логістичної регресії, що, в свою чергу, часто використовується для задач класифікації. Застосувати множинний дискримінантний аналіз для побудови, наприклад, багатовимірної поліноміальної регресії досить складно.

На рисунку 1.2 показано приклад застосування множинного дискримінантного аналізу та логістичної регресії для прогнозування успішності компаній у найближчі 5 років базуючись на їх фінансових показниках [21].

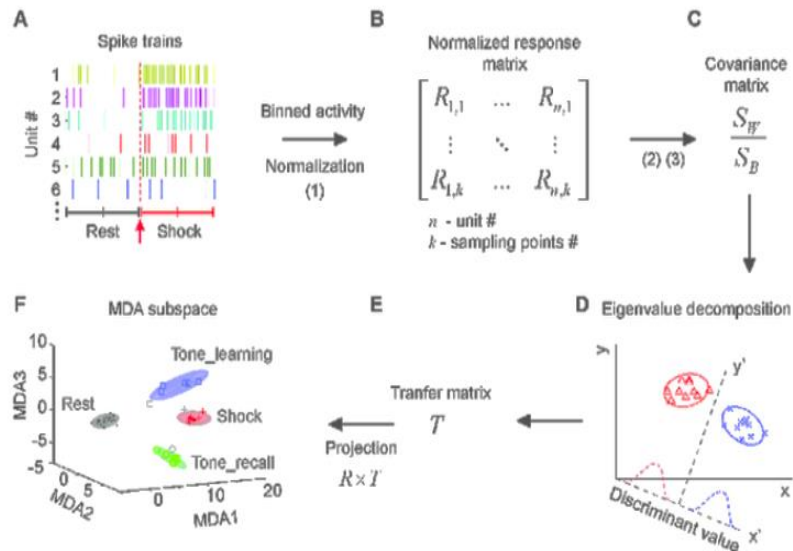


Рисунок 1.2 – Приклад застосування множинного дискримінантного аналізу та логістичної регресії [21]

1.2.3 Метод найменших квадратів

Метод найменших квадратів – це метод, що застосовується для побудови регресійних моделей таким чином, щоб якнайкраще відобразити залежність між вхідними незалежними змінними та вихідною залежною змінною [18, 19, 29].

Метод найменших квадратів ґрунтується на пошуку таких коефіцієнтів регресійної моделі, щоб мінімізувати різницю між прогнозованими значеннями вихідної змінної та значеннями, що отримані експериментально. Метод найменших квадратів використовується для пошуку коефіцієнтів одновимірної та багатовимірної поліноміальних регресійних моделей. Варто зауважити, що застосування методу найменших квадратів для пошуку коефіцієнтів багатовимірної регресійної моделі зводиться до трансформації останньої у кілька одновимірних регресій і подальшому розв'язку отриманого набору лінійних рівнянь. Приклад застосування методу найменших квадратів для побудови одновимірної лінійної регресійної моделі показано на рисунку 1.3.

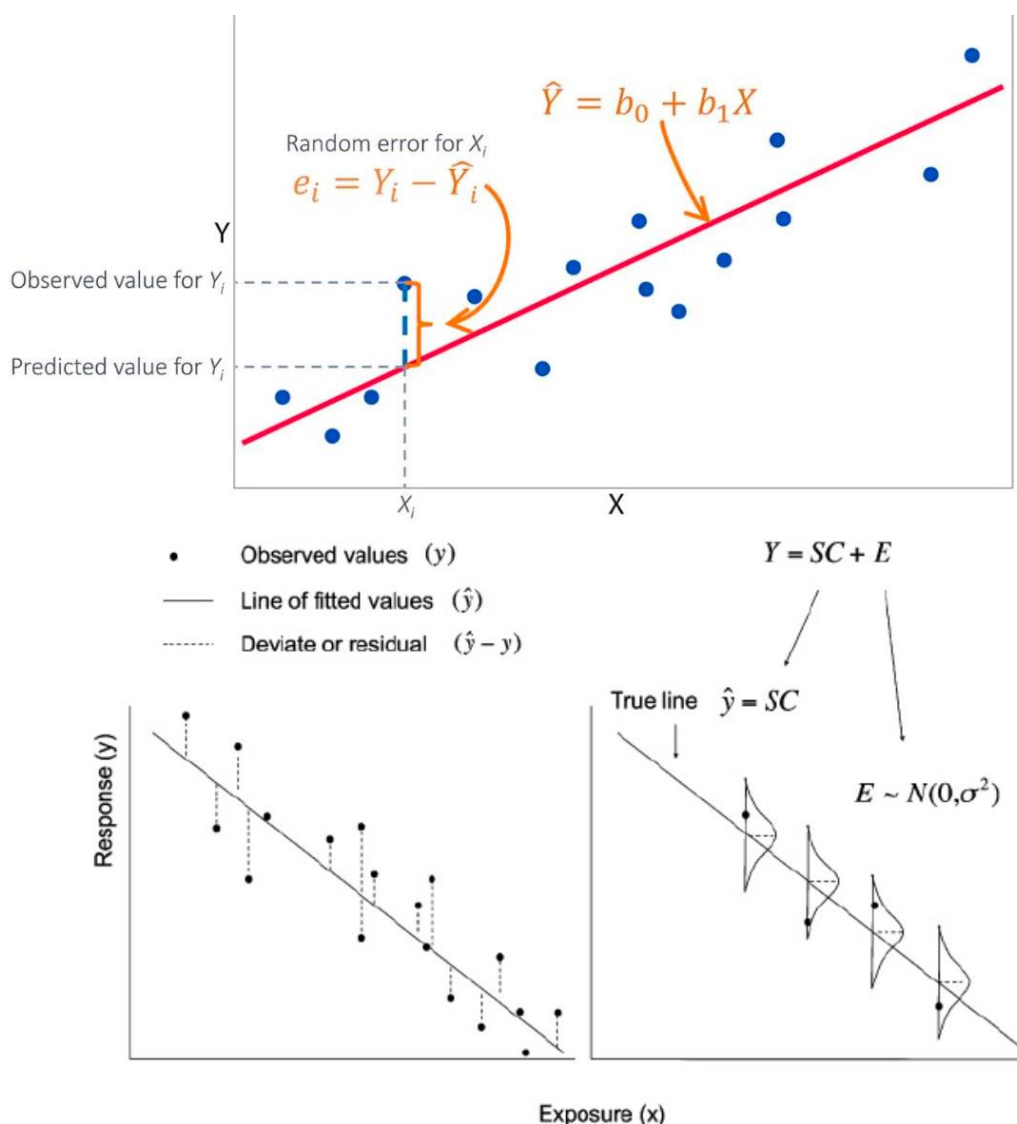


Рисунок 1.3 – Приклад застосування методу найменших квадратів для побудови одновимірної лінійної регресії [18]

1.2.4 Факторний аналіз

Факторний аналіз – це метод статистичного аналізу, що застосовується для ідентифікації взаємозв'язків між змінними та детекції непомітних на перший погляд структур або факторів. Даний метод є дуже поширеним у таких сферах, як економіка, маркетинг, соціологія для виявлення та демонстрації комплексних взаємозв'язків між змінними. Факторний аналіз може бути застосований для побудови регресійних моделей, зокрема для побудови багатовимірних поліноміальних регресій [1, 10, 23].

Факторний аналіз логічно та алгоритмічно розділяється на декілька етапів. На першому етапі відбувається збір експериментальних даних. Ці дані зберігаються у вигляді матриці, кожен стовпець якої відповідає конкретній змінній, а кожен рядок відображає одному експериментальному спостереженню. Після цього, за допомогою методу головних компонент чи факторного аналізу за власними числами, відбувається пошук прихованих факторів. Фактор – це приховане поєднання змінних, що разом впливають на значення вихідної змінної або задають йому тенденцію. Наступним кроком є інтерпретація та документація знайдених факторів. Далі, виявлені фактори застосовуються для побудови регресійних моделей.

1.2.5 Метод групового урахування аргументів

Метод групового урахування аргументів – це метод математичної статистика, що може використовуватися для побудови регресійних моделей. Суть цього методу полягає у тому, що змінні включають в модель певними підгрупами замість включення кожної змінної у модель окремо. Такий підхід призводить до ефективного врахування взаємозв'язків між змінним всередині підгрупи та їхнім впливом на кінцеве значення залежної змінної. Окрім того, об'єднання змінних у підгрупи дозволяє оптимізувати алгоритм з точки зору використання обчислювальних ресурсів, що особливо важливо у випадку необхідності опрацювання великого об'єму експериментальних даних [5, 8, 14].

Робота методу групового урахування аргументів логічно розділяється на декілька етапів. Спочатку, на основі певних критеріїв, змінні об'єднуються в підгрупи, що будуть враховуватись разом. Далі для кожного набору змінних відбувається побудова регресійної моделі (до уваги беруться тільки ті змінні, що належать конкретній підгрупі змінних). Таким чином буде отримано набір регресійних моделей, кількість яких рівна кількості підгруп вхідних змінних. На наступному етапі аналізується набір цих регресійних моделей і, за певними правилами, формуються регресійні моделі наступного рівня селекції. За допомогою перевірконої послідовності на кожному кроці селекції ідентифікується частковий опис з найменшою залишковою сумою квадратів.

Робота алгоритму зупиняється, коли залишкова сума квадратів для конкретного часткового опису стає вищою, ніж та, що отримана на попередньому етапі відбору. Цей частковий опис і буде результатом роботи методу групового урахування аргументів.

Найкращу ефективність даний метод показує тоді, коли є конкретні твердження чи припущення, що дозволяють об'єднати певні змінні у підгрупи. Проте, даний алгоритм має один суттєвий недолік – в процесі роботи алгоритму частина отриманих часткових розв'язків відкидається. Це може призвести до того, що раніше відкинутий частковий опис міг містити важливу ознаку, що могла значно покращити точність кінцевої моделі.

1.3 Критичний аналіз програмного забезпечення, що реалізує методи побудови багатовимірних регресій

Розглянемо існуючі програмні продукти, що реалізують різні методи побудови регресійних моделей. До таких програмних продуктів належать:

1.3.1 Scikit-learn

Scikit-learn [22] – це бібліотека для мови програмування Python, що об'єднує в собі різні засоби та методи для вирішення завдань за допомогою машинного навчання. Дану бібліотеку можна ефективно застосовувати для задач кластеризації, класифікації, обробки текстових даних, а також для побудови регресійних моделей. У собі бібліотека вміщує необхідні інструменти для знаходження коефіцієнтів одновимірних та багатовимірних регресій, а також інструменти, що дозволяють оцінити якість отриманих коефіцієнтів різними методами. Використання Scikit-learn значно полегшує процес застосування методів машинного навчання для вирішення прикладних задач. Це досягається за рахунок доступу до необхідних методів через зрозумілий та простий інтерфейс, який надає бібліотека Scikit-learn.

Перевагами даного програмного продукту однозначно є широкий набір доступних до використання інструментів і методів машинного навчання та лаконічний, зручний для використання та зрозумілий інтерфейс використання

цих методів. Однак, дана бібліотека погано пристосована для її використання у поєднанні з дуже великою кількістю даних. Крім того, цей програмний продукт не дозволяє гнучко керувати процесом побудови регресійних моделей.

1.3.2 NumPy

NumPy [17] – одна з ключових бібліотек Python, що застосовується для обчислень. Бібліотека об'єднує в собі велику кількість функцій для роботи з матрицями та масивами даних. В свою чергу, NumPy дуже часто використовується іншими бібліотеками, що потребують великої кількості математичних обчислень.

NumPy не містить в собі функціональності, що дозволяє напряму працювати з побудовою регресійних моделей, що значно ускладнює використання цього програмного продукту для вирішення таких завдань. Проте, велика кількість вже реалізованих математичних інструментів, дозволяє користувачу розробити будь-яку регресійну модель на базі цього програмного продукту. Такий підхід дозволяє керувати побудовою регресійних моделей набагато ефективніше, але значно збільшує складність знаходження роботи в процесі виконання таких завдань.

1.3.3 R-Stats

R-Stats [27] – статична бібліотека для аналізу даних у мові програмування R. Найчастіше застосовується для математичних обчислень, моделювання та аналізу даних. Дана бібліотека поєднує в собі велику кількість функцій, що значно розширюють потенціал R для статичного аналізу даних.

Бібліотека R-Stats об'єднує в собі функції для статичного аналізу, прогнозування, візуалізації певних даних, а також елементарних математичних операцій. Нажаль, R-Stats не пропонує користувачу зручного інтерфейсу для функцій, що можуть бути застосовані для побудови регресійних моделей. Такий недолік призводить до того, що побудова регресійних моделей з використанням таких технологій стає складною і нетривіальною задачею, а процес гнучкого налаштування процесу побудови моделі регресії стає майже неможливим.

1.3.4 STATISTICA

STATISTICA [25] – програма для всебічного і різностороннього статистичного аналізу даних. Розробкою даного продукту займається компанія StatSoft. У собі STATISTICA об'єднує безліч функцій, що можуть використовуватися для вирішення різноманітних задач, таких як аналіз даних, а також візуалізація, добування та управління статичнісними даними.

Бібліотека STATISCTICA підтримує можливість побудови різних регресійних моделей. За допомогою даного програмного продукту легко виконуються задачі побудови одновимірних та багатовимірних поліноміальних регресій. Наприклад, завдання побудови багатовимірної поліноміальної регресії виконується за допомогою функції Polynomial Regression. Дана функція самостійно ідентифікує конфігурацію багатовимірної регресії в процесі роботи, тому гнучка конфігурація побудови моделі для користувача не завжди є можливою, що стає недоліком даного продукту при його використанні для вирішення подібних завдань.

1.3.5 SPSS Statistics

SPSS Statistics [24] – програмний продукт, що застосовується для статистичного аналізу даних. За допомогою даного програмного забезпечення також можна з легкістю виконувати задачі одновимірних та багатовимірних поліноміальних регресій. Загалом, підтримуються всі поширені види регресійних моделей. Головний підхід роботи даного програмного проодукту при побудові регресійних моделей – визначення значимості членів регресійної моделі.

Як і STATISTICA, SPSS Statistics бере на себе відповідальність визначення конфігурації побудованої регресійної моделі, що стає недоліком даного програмного продукту тоді, коли користувачу потрібна можливість гнучко налаштовувати процес побудови регресійної моделі.

1.4 Постановка задачі

Проведений критичний аналіз методів побудови багатовимірної регресії показав, що всі розглянуті методи дають прийнятний результат свого

застосування тільки при наявності великої кількості експериментальних даних. Жоден з розглянутих методів не заточений під роботу з малим об'ємом експериментальних даних. При цьому, збір великої кількості експериментальних даних деколи є неможливим.

Тому, метою даної роботи є підвищення ефективності побудови багатовимірної регресії на малому об'ємі експериментальних даних. Для досягнення мети необхідно вирішити наступні задачі:

- розробка методології проведення статистичного імітаційного експерименту дослідження ефективності методу побудови багатовимірної лінійної регресії, заданої надлишковим описом, по невеликому об'єму експериментальних даних;
- розробка та обґрунтування архітектури програмного забезпечення, що реалізує оригінальний метод побудови багатовимірних лінійних регресій, заданих надлишковим описом, та системи статистичного імітаційного моделювання, що дозволяє знайти область його ефективного застосування;
- розробка програмного забезпечення, що реалізує оригінальний метод побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних та систему статистичного імітаційного моделювання його ефективності;
- розробка рекомендацій по використанню досліджуваного методу на основі проведених статистичних досліджень.

Створена бібліотека повинна відповідати наступним вимогам:

- реалізує оригінальний алгоритм для побудови багатовимірної регресії на малому об'ємі експериментальних даних;
- кросплатформенність (Windows, Linux, MacOS);
- надає користувачу інтуїтивно зрозумілий інтерфейс для використання розробленого алгоритму для побудови багатовимірної регресії на основі малого об'єму експериментальних даних;
- простота і зручність у інтеграції та використанні.

Висновки до розділу

У першому розділі проведено аналіз предметної області, де описано суть регресійного аналізу та наведено математичні моделі одновимірних гіперболічної та поліноміальної регресій, а також багатовимірних логістичної та поліноміальної регресій. Далі було проведено аналіз існуючих методів побудови регресійних моделей та визначено сфери їх застосування. Було розглянуто метод максимальної правдоподібності, множинний дискримінаційний аналіз, метод найменших квадратів, факторний аналіз та метод групового урахування аргументів. Також були проаналізовані основні сучасні програмні продукти, що використовують розглянуті методи (Scikit-learn, NumPy, R-Stats, STATISTICA, SPSS Statistics), визначено основні переваги та недоліки даних програмних продуктів. Результати проведеного аналізу показують, що жоден з розглянутих методів та програмних продуктів не фокусується на побудові регресійних моделей саме на малому об'ємі експериментальних даних. Було визначено мету даної роботи, що полягає у підвищенні ефективності побудови багатовимірних поліноміальних регресій на малому об'ємі експериментальних даних. Було чітко визначено завдання, виконання яких допоможе досягти поставленої мети.

2 ОРИГІНАЛЬНИЙ МЕТОД ПОБУДОВИ БАГАТОВИМІРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ НА МАЛОМУ ОБ'ЄМІ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

Мета методу, що досліджується, полягає в підвищенні ефективності знаходження оцінок коефіцієнтів БЛР, заданої надлишковим описом, загальною алгоритмічною процедурою метода найменших квадратів (МНК), що задається загальною формулою:

$$\hat{b} = (A^T A)^{-1} A^T Y, \quad (2.1)$$

де невідомі коефіцієнти вектора b в структуру БЛР входять лінійно.

2.1 Опис методу підвищення продуктивності загальної процедури МНК

2.1.1 Постановка задачі

Модель БЛР, заданої надлишковим описом[30]:

$$Y(\bar{x}) = b_0 + \sum_{j=1}^m b_j x_j + E, \quad (2.2)$$

$ME = 0, DE < \infty$. Деякі коефіцієнти можуть бути нулями.

Задача. За результатами статистичного випробовування ($\bar{x}_i \rightarrow y_i, i = \overline{1, n}$) оцінити невідомі коефіцієнти $b_j, j = \overline{1, m}$. В модель задається фіктивна вхідна змінна x_{m+1} , у якої $b_{m+1} = 0$ [30].

2.1.2 Етапи методу

Перший етап. Загальною процедурою МНК знаходяться оцінки $\bar{b}_j, j = \overline{0, m+1}$ за результатами імітації активного експерименту[30].

Другий етап. Всі коефіцієнти $\bar{b}_j, j = \overline{0, m+1}$ розбиваються на два класи M_1 та M_2 наступним алгоритмом[30]:

Перший крок. Рангування оцінок за моделями їх значень: $|\bar{b}_{j_1}| \geq \dots \geq |\bar{b}_{j_{m+1}}|$ [30].

Другий крок. Коефіцієнт $\bar{b}_{j_{m+1}}$ відноситься до класу M_2 , а коефіцієнти \bar{b}_{j_1} та \bar{b}_0 відносяться до класу M_1 [30].

l -тий крок. Коефіцієнт \bar{b}_{j_l} відноситься до класу M_1 [30]. Якщо:

$$\frac{1}{l} \sum_{i=1}^l |\bar{b}_{j_i}| - |\bar{b}_{j_{l+1}}| < |\bar{b}_{j_{l+1}}| - |\bar{b}_{j_{m+1}}|, \quad (2.3)$$

то $\bar{b}_{j_{l+1}}$ відноситься до класу M_1 . [30] В протилежному випадку розбиття вважається завершеним і матиме наступний вигляд:

$$M_1 = \{\bar{b}_{j_1} \dots \bar{b}_{j_l}\}, M_2 = \{\bar{b}_{j_{l+1}} \dots \bar{b}_{j_{m+1}}\} \quad (2.4)$$

В результаті остаточного розбиття $\bar{b}_{j_{m+1}}$ повинен належати M_2 .

Третій етап. З моделі (1) виключаються коефіцієнти з класу M_2 . Оцінки коефіцієнтів з класу M_1 знаходяться за локальною процедурою МНК[30].

Четвертий етап. До отриманої моделі БЛР реалізуються етапи 1-3[30].

Кінцевий етап. Описана процедура продовжується доти, поки M_1 не буде складатися лише з одного коефіцієнта при фіктивній вхідній змінній. Оцінка коефіцієнтів знайденої структури БЛР знаходиться загальною процедурою МНК[30].

2.2 Методологія статистичного дослідження ефективності оригінального методу побудови БЛР, заданої надлишковим описом

Під час статистичного дослідження задаються різні розподіли випадкової величини E – нормальний; рівномірний на відрізку $[-c, c]$, ($c > 0$); сума нормального та рівномірного розподілу.

Функція щільності ймовірності для рівномірного розподілу на відрізку $[-c, c]$ визначається наступним чином:

$$f(x) = \frac{1}{2c}, \text{ для } -c \leq x \leq c,$$

де $f(x)$ – значення щільності ймовірності в точці x , а $[-c, c]$ – інтервал, на якому визначається розподіл.

Функція щільності нормального розподілу має вигляд:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (2.5)$$

де $f(x, \mu, \sigma)$ – значення щільності ймовірності в точці x , μ – математичне очікування, а σ^2 – дисперсія нормального розподілу.

2.2.1 Параметри моделювання, що змінюються в процесі статистичних досліджень:

- функція щільності розподілу E ;
- σ^2 дисперсія нормального розподілу, математичне очікування завжди тотожне 0;
- $c > 0$, параметр рівномірного розподілу;
- $\min |b_j| > 0$, починаючи з величини 0,5;
- m – параметр розмірності БЛР;
- кількість нульових коефіцієнтів БЛР, що задається при імітації активного експерименту;
- значення ненульових коефіцієнтів БЛР, що задається при імітації активного експерименту;
- кількість імітацій випробовувань при фіксованих значеннях параметрів;
- кількість повторів статистичного імітаційного експерименту.

2.2.2 Процедура статистичного імітаційного моделювання

Перший етап. Фіксуються значення параметрів моделювання і реалізується задана кількість імітаційних випробовувань.

Другий етап . По результатам імітаційних випробовувань з використанням загальної процедури МНК, знаходяться наступні середні характеристики:

- відсоток знайдених правильних структур БЛР;
- середнє значення інтегральної похибки компонент вектора оцінок невідомих коефіцієнтів від їх значень, заданих в імітації, по правильно знайденим структурам БЛР. Інтегральна кількісна міра задається виразом:

$$\|\hat{b} - b\| = \left\| \frac{\hat{b}}{\|\hat{b}\|} - \frac{b}{\|b\|} \right\|,$$

де $\|x\| = \sqrt{\sum_{i=1}^k x_i^2}$, k – розмірність вектора x .

В алгоритмі імітаційного статистичного моделювання перший і другий етап повторюються задану кількість раз.

Останній етап. Аналіз результатів статистичного імітаційного моделювання, обґрунтування області ефективного використання методу побудови БЛР, заданої надлишковим описом, по невеликій кількості експериментальних даних.

Висновки до розділу

В даному розділі описано оригінальний метод побудови багатовимірних лінійних регресій, заданих надлишковим описом. Завдяки застосуванню описаного методу покращення ефективності роботи методу найменших квадратів можна досягти значного успіху при побудові багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних. Також проведено опис методології статистичного дослідження ефективності описаного методу побудови БЛР, заданої надлишковим описом, на малому об'ємі експериментальних даних.

3 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

3.1 Опис вимог до програмного забезпечення

По перше, розроблене програмне забезпечення повинно ефективно реалізовувати досліджуваний метод побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних. Пропонується, що дана частина програмного забезпечення буде реалізована як окрема бібліотека, яку користувач зможе достатньо просто використати при розробці свого програмного забезпечення для побудови багатовимірної лінійної регресії по малому об'єму експериментальних даних. На даний момент найпопулярнішою і найпростішою мовою програмування, що дозволяє ефективно вирішувати завдання побудови багатовимірної лінійної регресії, є мова програмування Python. Тому, бібліотека, що реалізує оригінальний алгоритм побудови багатовимірної лінійної регресії по малому об'єму експериментальних даних, повинна бути реалізована саме цією мовою програмування.

По друге, розроблене програмне забезпечення повинно реалізовувати розроблену методологію статистичного дослідження ефективності оригінального методу побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних. Ця частина програмного забезпечення буде реалізована як застосунок з графічним інтерфейсом користувача. Система статистичного дослідження ефективності оригінального методу побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних повинна дозволити користувачу ввести необхідні параметри моделювання, що передбачені розробленою методологією статистичного дослідження, та, на їх основі, виконати розроблену процедуру статистичного імітаційного моделювання ефективності оригінального методу побудови БЛР на малому об'ємі експериментальних даних. Застосунок має бути написаний мовою програмування Python. Це дозволить ефективно використовувати розроблену бібліотеку в межах цієї системи.

Розроблена система статистичного дослідження ефективності оригінального методу побудови БЛР повинна надавати користувачеві можливість вводу наступних параметрів:

- розмірність БЛР (від 1 до 50);
- кількість нульових коефіцієнтів БЛР;
- значення ненульових коефіцієнтів;
- кількість імітацій випробувань при фіксованих значеннях параметрів;
- кількість повторів експерименту при фіксованих значеннях параметрів;
- тип розподілу випадкової величини E та його параметрів (математичне очікування та дисперсія).

Далі застосунок повинен провести валідацію введених користувачем даних. Якщо при цьому виникли певні помилки, графічний інтерфейс повинен відобразити їх користувачеві. У випадку успішної валідації вхідних параметрів, застосунок повинен виконати розроблену процедуру статистичного імітаційного моделювання ефективності оригінального методу побудови БЛР.

3.2 Розробка архітектури програмного забезпечення

Існує велика кількість різних типів архітектури програмного забезпечення, серед яких найбільш відомими є монолітна, мікросервісна, клієнт-серверна та подійно орієнтована архітектури[31].

Перевагами монолітної архітектури є просте розгортання ПЗ, найкраща продуктивність викликів компонентів в межах системи, достатньо простий процес тестування та відлагоджування ПЗ[31]. Недоліками даного типу архітектури ПЗ є погана масштабованість, сильні залежності між частинами програмного забезпечення та велика складність внесення змін в уже робоче ПЗ[31].

Перевагами мікросервісної архітектури є гнучкість в розгортанні та масштабуванні ПЗ, незалежність частин ПЗ одне від одного, проста процедура

внесення змін[31]. Недоліками даного типу архітектури ПЗ є складність управління міжсервісною комунікацією, адже потрібні додаткові ресурси на відстеження активності та моніторинг роботи сервісів[31].

Клієнт-серверна архітектура дозволяє легше забезпечити безпеку даних, а також зменшує навантаження на стороні клієнта[31]. Проте, суттєвим недоліком є залежність від мережі, а також нестабільна робота при високому навантаженні сервера[31].

Подійно-орієнтована архітектура дозволяє значно легше впроваджувати асинхронну взаємодію, але може мати високу складність відлагодження та забезпечення зв'язності даних[31].

За результатами аналізу різних типів архітектури програмного забезпечення, для імплементації бібліотеки, що реалізує оригінальний метод побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних, а також системи статистичного імітаційного моделювання ефективності використання цього методу, було обрано монолітну архітектуру, оскільки її переваги є абсолютно необхідними в цій системі (висока продуктивність викликів), а недоліки не виявляються через невеликий розмір ПЗ. На рисунку 3.1 представлено розроблену архітектуру програмного забезпечення.

Відповідного до того, як відображено на діаграмі, система статистичного імітаційного моделювання ефективності досліджуваного методу (моноліт) вміщує в собі частини ПЗ, що відповідають за:

- графічний інтерфейс користувача;
- імітацію активних експериментів;
- пошук коефіцієнтів БЛР (з використанням розробленої бібліотеки);
- обчислення метрик для оцінки ефективності знайденого рішення.

Бібліотека, що реалізує оригінальний алгоритм побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних, в свою чергу, вміщує частини ПЗ, що відповідають за:

- застосування загальної процедури МНК для пошуку початкових, проміжних і фінальних оцінок коефіцієнтів БЛР;

- розподіл коефіцієнтів БЛР на два класи: ті, що впливають на значення вихідної змінної, і ті, що не впливають;
- ітеративний процес застосування МНК та розподілу коефіцієнтів БЛР.

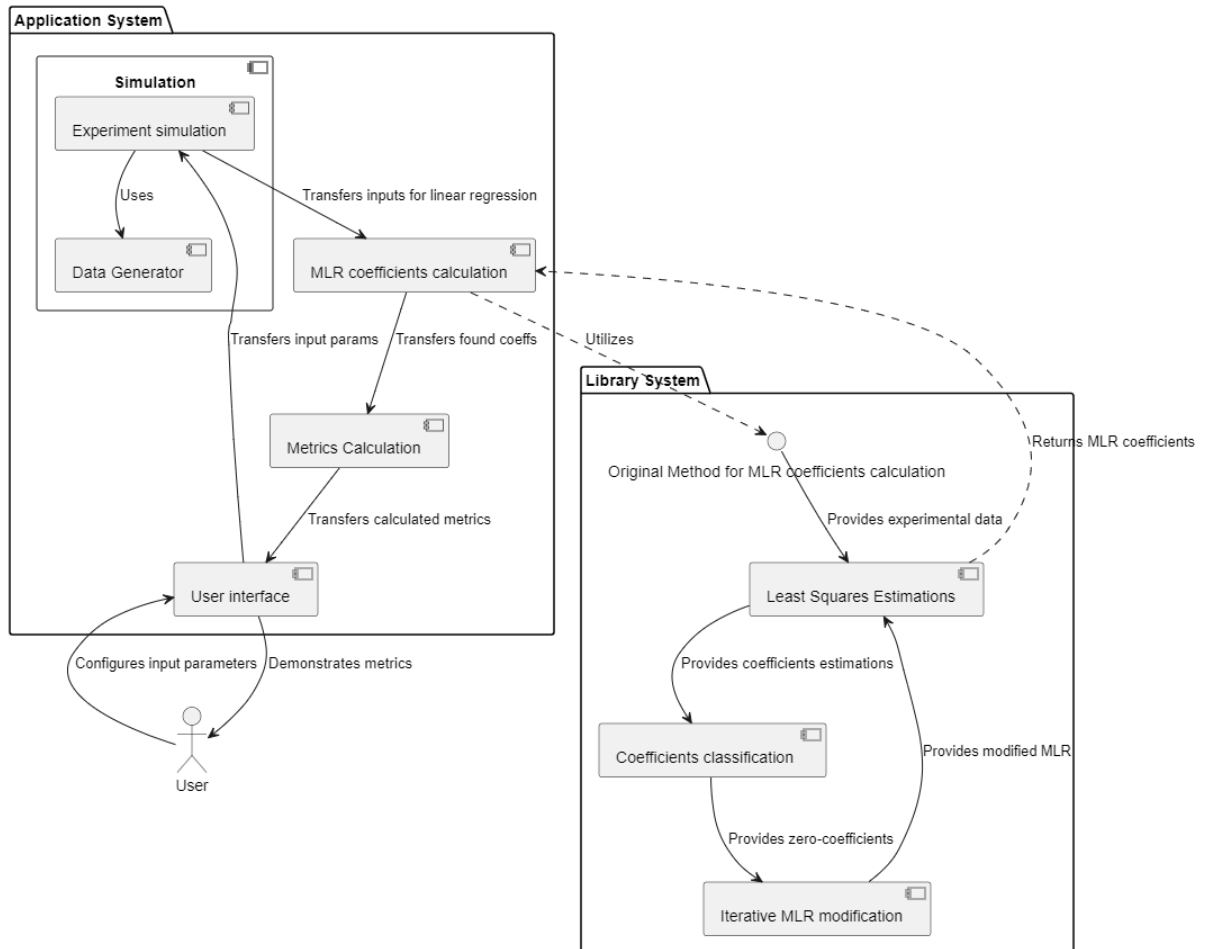


Рисунок 3.1 – Діаграма архітектури програмного забезпечення

3.3 Опис програмного забезпечення

Для розробки діаграм класів ПЗ використовувалоя розширення PlantUML для Visual Studio Code.

Діаграма класів бібліотеки, що реалізує оригінальний алгоритм побудови БЛР на малому об'ємі експериментальних даних, представлено на рисунку 3.2.

Отже, бібліотека складається з двох класів: клас Matrix (для зручного доступу до даних), та клас MultivariableLinearRegressionModel, що реалізує оригінальний алгоритм побудови БЛР заданої надлишковим описом на малому об'ємі експериментальних даних.

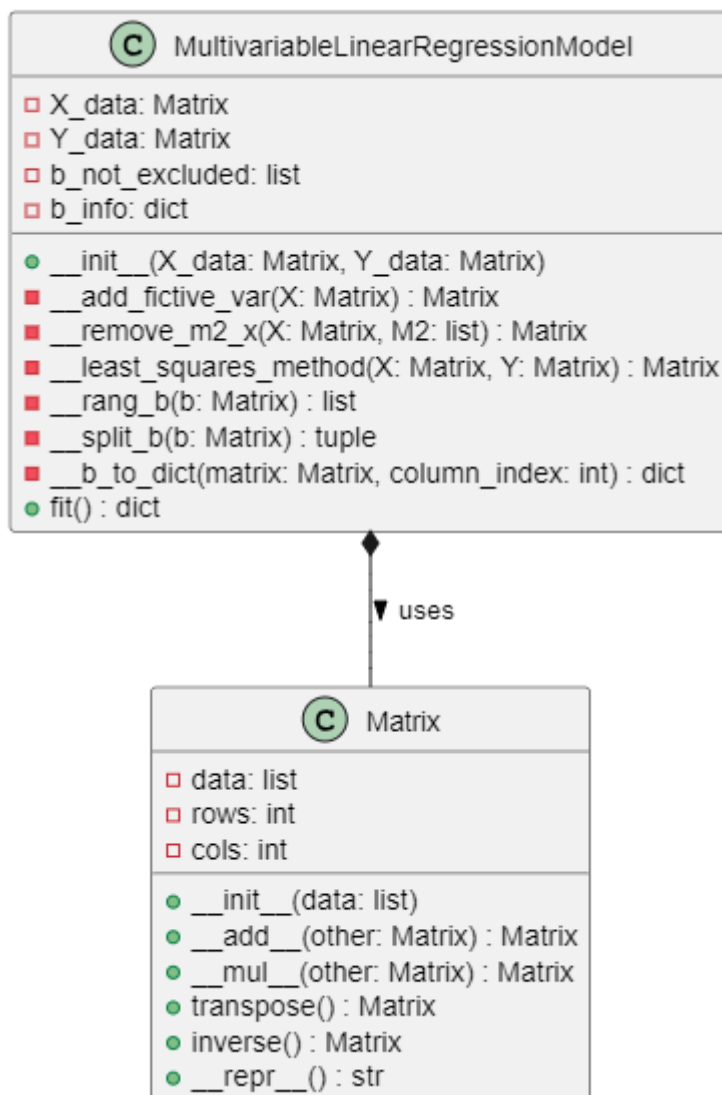


Рисунок 3.2 – Діаграма класів бібліотеки, що реалізує досліджуваний метод побудови БЛР, заданої надлишковим описом

Клас `Matrix` реалізує зручну роботу з матрицями. Використовується в межах розробленої бібліотеки та в межах розробленого застосунку. Може використовуватись користувачем бібліотеки для організації коректної роботи з матрицями в клієнтському застосунку. Клас `Matrix` зберігає в собі поле `data`, що є списком списків (представляє матрицю), а також поля `rows` та `cols`, що визначають розмірність матриці та використовують для коректного доступу до елементів матриці. Конструктор класу `Matrix` дозволяє створити об'єкт, передавши йому дані матриці. Розмірність матриці буде визначено автоматично.

Клас `Matrix` містить наступні методи:

- `__add__` – реалізує операцію додавання матриць;
- `__mul__` – реалізує операцію множення матриць;
- `transpose` – повертає транспоновану матрицю;
- `inverse` – знаходить обернену матрицю та повертає її;
- `__repr__` - реалізує логіку відображення матриці як рядка.

Клас `MultivariableLinearRegressionModel` реалізує досліджуваний метод. Конструктор цього класу приймає два параметри – матрицю X (значення незалежних вхідних змінних) та матрицю Y (значення вихідної змінної). Окрім цього, `MultivariableLinearRegressionModel` містить наступні поля:

- `b_not_excluded` – містить список невиключених коефіцієнтів БЛР;
- `b_info` – містить оцінки коефіцієнтів БЛР.

Клас `MultivariableLinearRegressionModel` містить наступні методи:

- `__add_fictive_var` – додає фіктивну вхідну змінну в БЛР;
- `__remove_m2_x` – видаляє з БЛР коефіцієнти, що віднесені до класу `M2`, тобто вважаються нулями;
- `__least_squares_method` – реалізує загальну процедуру МНК (використовується для пошуку початкових, проміжних та фінальних оцінок коефіцієнтів БЛР);
- `__rang_b` – реалізує логіку рангування коефіцієнтів БЛР за модулями їх значень;
- `__split_b` – реалізує логіку поділу коефіцієнтів БЛР на два класи за їх рангами;
- `__b_to_dict` – перетворює вектор оцінок коефіцієнтів БЛР у словник, де ключем є індекс коефіцієнта БЛР, а значенням – його оцінка;
- `fit` – реалізує логіку роботи досліджуваного методу для пошуку ефективних оцінок коефіцієнтів БЛР.

Таким чином, розроблена бібліотека містить всі необхідні засоби для ефективного використання досліджуваного методу користувачем для вирішення своїх задач з побудови БЛР на малому об'ємі експериментальних даних.

Діаграма класів застосунку, що реалізує систему статистичного імітаційного моделювання ефективності досліджуваного методу побудови БЛР, заданої надлишковим описом, на малому об'ємі експериментальних даних, представлена на рисунку 3.3.

Взаємодія користувача з застосунком починається із графічного інтерфейсу – користувачу необхідно ввести вхідні дані. Далі ці параметри будуть передані до класу MainApp, що відповідає за всю роботу застосунку. Цей клас містить один метод – run_simulation, що відповідає за правильну обробку вхідних даних, виклики всіх інших компонентів застосунку з коректними параметрами для отримання результатів моделювання. Таким чином, спочатку викликається ExperimentSimulator, і за допомогою його методу simulate, отримуються дані експерименту, а саме матриця вхідних змінних та матриця зі значеннями вихідної змінної. Далі отримані дані експерименту передаються RegressionModel і викликається метод fit, що використовує розроблену бібліотеку, що реалізує досліджуваний метод побудови БЛР. Результатом роботи цього методу буде RegressionResults, що містить в собі актуальні значення коефіцієнтів БЛР, значення оцінок коефіцієнтів БЛР, що знайдені загальною процедурою метода найменших квадратів, а також значення коефіцієнтів БЛР, що знайдені за допомогою досліджуваного методу. Далі ці результати передаються в RegressionResultsAnalyzer, що відповідає за обчислення відповідних метрик на основі цих результатів (відсоток правильно знайдених структур БЛР, інтегральна похибка вектору оцінок, знайдених за допомогою МНК на досліджуваного методу, а також відсоток покращення ефективності оцінок коефіцієнтів БЛР завдяки використанню досліджуваного методу).

В результаті, отримані метрики будуть відображені за допомогою графічного інтерфейсу користувача.

Таким чином, застосунок поєднує в собі класи для відображення графічного інтерфейсу, генерації даних активного експерименту, використання розробленої бібліотеки та аналізу отриманих оцінок коефіцієнтів БЛР. Застосунок повністю реалізує розроблену процедуру статистичного імітаційного

моделювання ефективності застосування досліджуваного метода побудови БЛР, заданої надлишковим описом, на малому об'ємі експериментальних даних.

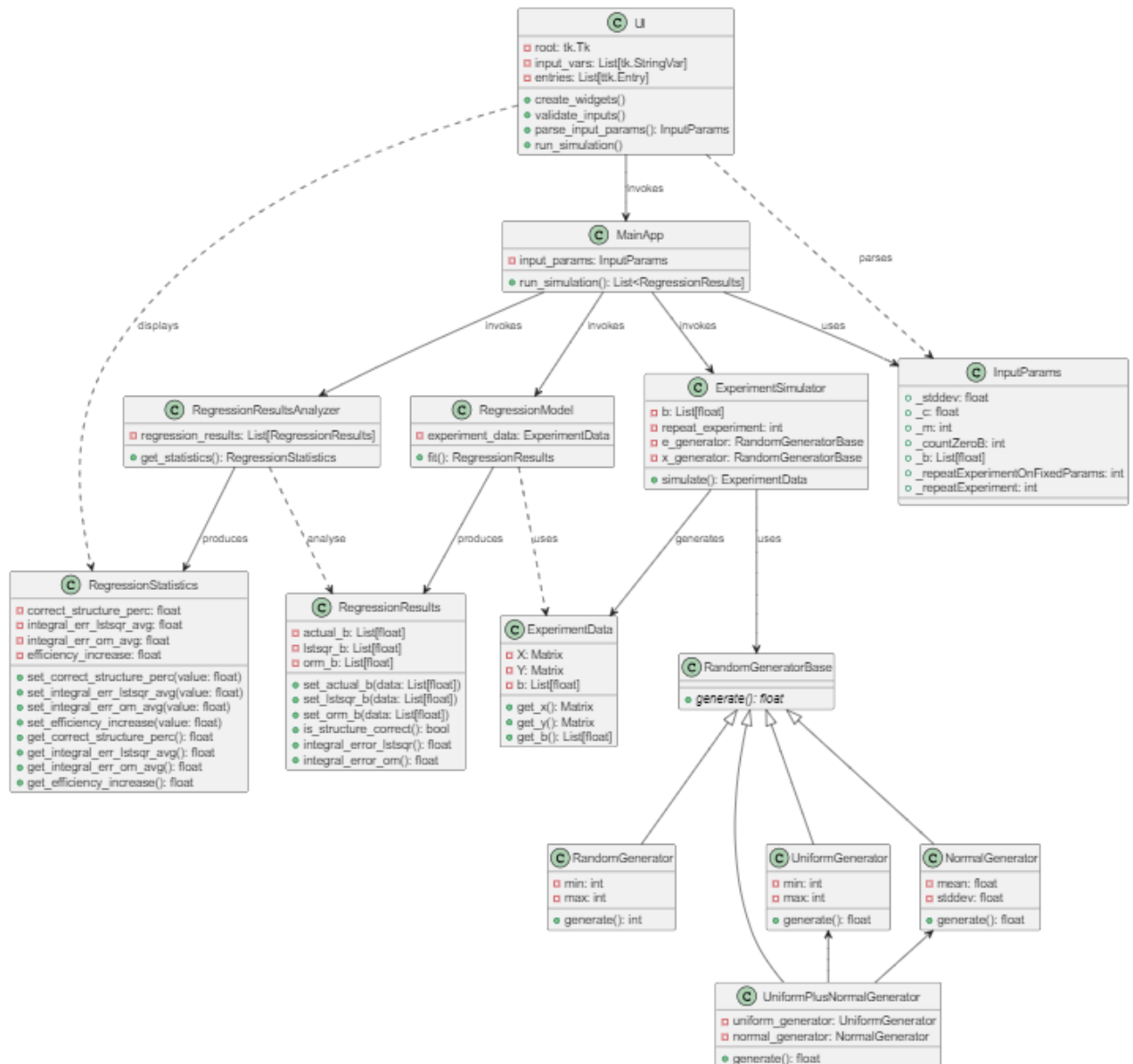


Рисунок 3.3 – Діаграма класів системи статистичного імітаційного моделювання ефективності досліджуваного метода побудови БЛР, заданої надлишковим описом, на малому об'ємі експериментальних даних

3.4 Графічний інтерфейс користувача

Графічний інтерфейс користувача на початковому етапі роботи застосунку виглядає так, як зображено на рисунку 3.4.

Користувачу доступні такі параметри для введення, як розмірність БЛР (за замовчуванням значення 10), кількість нульових коефіцієнтів (за замовчуванням 5), кількість повторів в одному експерименті (за замовчуванням 20), кількість

експериментів (за замовчуванням 10), тип та параметри розподілу випадкової величини E (за замовчуванням рівномірний розподіл від -2 до 2) і значення коефіцієнтів БЛР (за замовчуванням всі нулі).

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
0	0	0	0	0	0	0	0	0	0	0

Рисунок 3.4 – Початковий вигляд графічного інтерфейсу користувача

Щоб задавати конкретний тип розподілу випадкової величини E , користувач може використовувати відповідний перемикач. Користувачу буде запропоновано ввести такі параметри, що відповідають вибраному розподілу. Наприклад, на рисунку 3.5 відображено ситуацію, коли обрано нормальний розподіл для випадкової величини E .

На рисунку 3.6 відображено ситуацію, коли для випадкової величини E обрано суму нормального та рівномірного розподілу. В такому випадку користувач повинен ввести і стандартне відхилення для нормального розподілу, і параметр рівномірного розподілу.

На рисунку 3.7 відображено ситуацію, коли користувач ввів від'ємне значення стандартного відхилення нормального розподілу випадкової величини

Е. Спочатку поле вводу виділяється червоним, як некоректне значення, а при натисканні на кнопку “Run experiments”, користувачу повідомляється про помилку у введених даних. Виділення червоним кольором неправильно введених даних працює для всіх полів вводу.

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Normal distributions standart deviation (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
0	0	0	0	0	0	0	0	0	0	0

Рисунок 3.5 – Обрано нормальний розподіл для випадкової величини E

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

Normal distributions standart deviation (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
0	0	0	0	0	0	0	0	0	0	0

Рисунок 3.6 – Обрано суму нормального та рівномірного розподілу для випадкової величини E

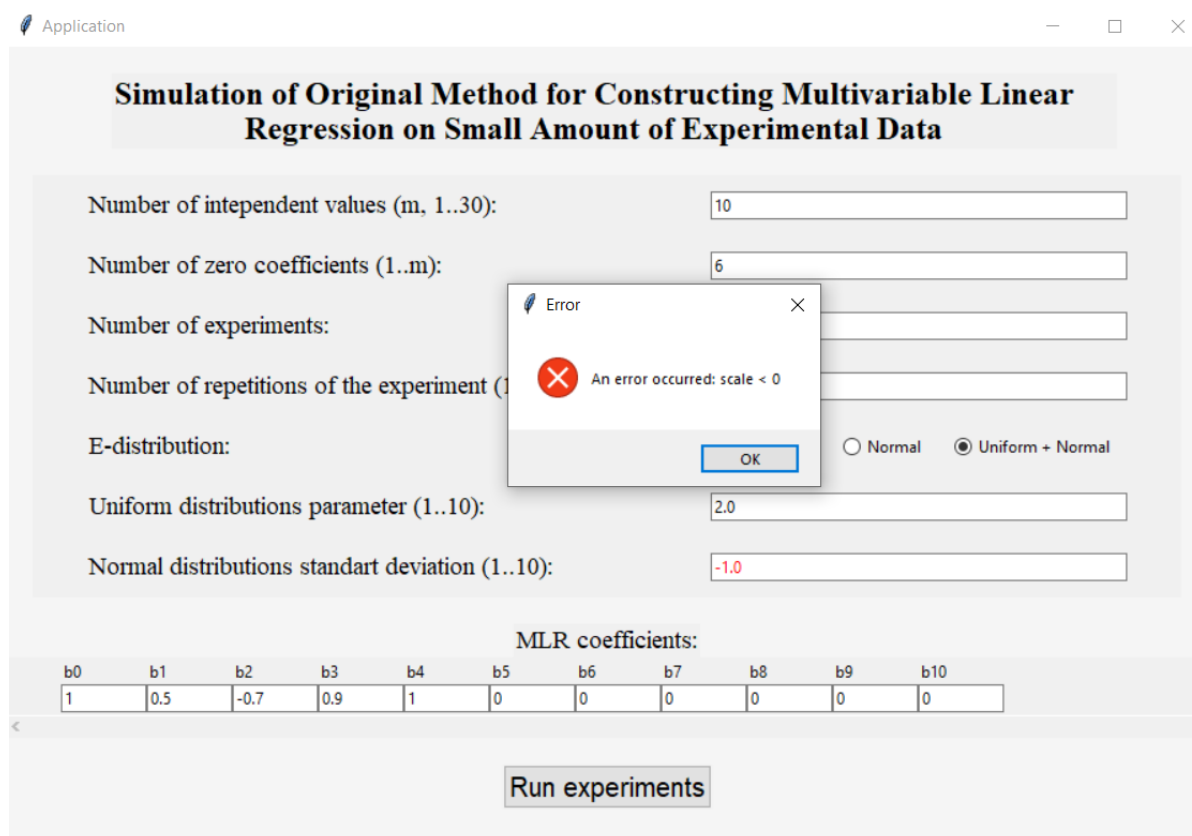


Рисунок 3.7 – Користувач ввів некоректне значення параметра нормального розподілу випадкової величини E

Також передбачено функціональність інформування користувача про кількість введених ненульових значень коефіцієнтів БЛР. Якщо користувач ввів рівно стільки ненульових коефіцієнтів, скільки потрібно, то коефіцієнти будуть підсвічуватись чорним кольором (рисунок 3.8). Якщо кількість введених ненульових коефіцієнтів менша (рисунок 3.9) або більша (рисунок 3.10), ніж потрібно, то поля вводу коефіцієнтів буду видулені синім кольором. Теоретично, це не є критичною помилкою, і експеримент можна буде успішно запустити в даному випадку, тому таке інформування користувача є чисто інформатичним.

Кількість полів для вводу коефіцієнтів динамічно змінюється в залежності від введеного користувачем значення розмірності БЛР (рисунок 3.11). Якщо користувач введе таку кількість вхідних змінних, що поля для їх вводу не помістяться в межах вікна, то буде створено смугу для прокручування, що дозволить користувачу прокручувати зону введення значень коефіцієнтів БЛР (рисунок 3.12).

Application - □ ×

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
1	0.5	-0.7	0.9	1	1	0	0	0	0	0

Рисунок 3.8 – Введено коректну кількість ненульових коефіцієнтів

Application - □ ×

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
1	0.5	-0.7	0.9	1	0	0	0	0	0	0

Рисунок 3.9 – Введено менше ненульових коефіцієнтів, ніж очікується

Application

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
1	0.5	-0.7	0.9	1	0	1	0	1	0	0

Run experiments

Рисунок 3.10 – Введено більше ненульових коефіцієнтів, ніж очікується

Application

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5
0	1	0	3	0	0

Run experiments

Рисунок 3.11 – Кількість незалежних змінних задано як 5, користувачу надано 6 полів для вводу значень коефіцієнтів БЛР (5 незалежних змінних + вільний член)

Application

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

MLR coefficients:

b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	b14	b15
0	0	0	0	0	0	0	0	0	0	0	0	0

Рисунок 3.12 – Кількість незалежних змінних задано як 25, з'явилась смуга для прокручування, що надає змогу задати значення всіх необхідних коефіцієнтів

Application

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
1	2	-1	2	1	0	0	0	0	0	0

Experiment results:

Percentage of correctly found MLR structures	100.00%
Average integral error while using least squares	0.526
Average integral error while using original method	0.215
Percent improvement in least squares efficiency	59.20%

Рисунок 3.13 – Результати експерименту

На рисунку 3.13 відображено результати роботи системи статистичного імітаційного моделювання ефективності застосування досліджуваного методу. Таким чином, для БЛР, розмірністю 10, при кількості нульових коефіцієнтів 6,

20 випробуваннях в одному експерименті, при 10 повторах експерименту, при параметрів рівномірного розподілу 2, отримано 100% точності пошуку правильної структури БЛР і покращення ефективності оцінок коефіцієнтів БЛР на 59%. Далі, параметр рівномірного розподілу було змінено на 10 (рисунок 3.14), отримані результати моделювання кажуть, що правильну структуру БЛР було знайдено у 50% спроб, а покращення ефективності оцінок на правильно знайдених структурах БЛР складає 42%. Зменшення частки правильно знайдених структур БЛР тут очікувано, оскільки було збільшено параметр рівномірного розподілу випадкової величини E. Варто відмітити, що значення незалежних вхідних змінних кожного разу генеруються заново випадковим чином у діапазоні від 1 до 20.

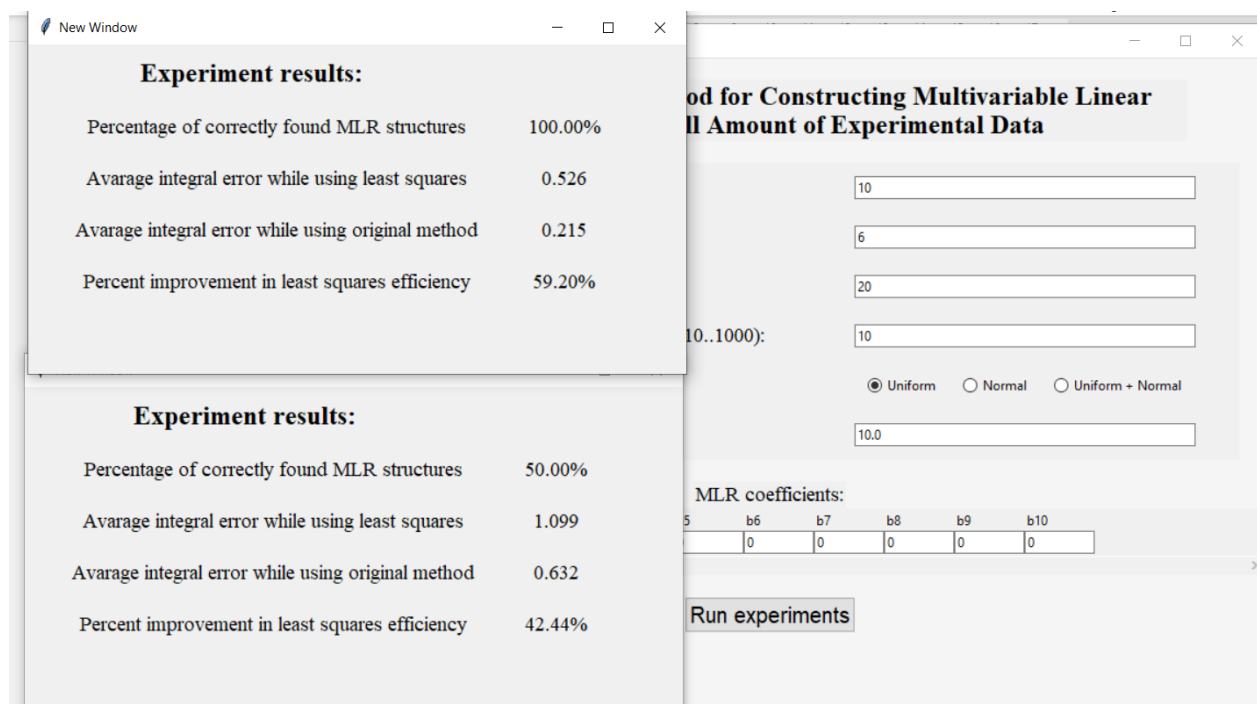


Рисунок 3.14 – Користувач змінив параметр рівномірного розподілу величини E і тепер може порівняти результати роботи застосунку

3.5 Інструкція з розгортання програмного забезпечення

Користувач може використовувати як і розроблену бібліотеку самостійно, так і розроблену систему статистичного імітаційного моделювання ефективності досліджуваного методу разом з бібліотекою. При будь-якому сценарії використання першим етапом буде розгортання бібліотеки, щоб зробити її доступною для використання. У даному випадку рекомендується локальне розгортання бібліотеки для її включення до цільового програмного забезпечення. Розроблена бібліотека вже містить в собі заготовлені файли, що дозволяють дуже швидко розгорнути без необхідності розбиратись у коді чи змінювати його. Для початку користувач повинен клонувати репозиторій із кодом бібліотеки. Далі, з папки, де збережено код бібліотеки, користувач повинен запустити команду `python setup.py bdist_wheel`. За допомогою виконання цієї команди буде проведено збірку бібліотеки. Результат роботи цієї команди предсталено на рисунку 3.15. В результаті буде створено відподвійний whl файл. Наступним кроком буде встановлення бібліотеки. Для цього користувач повинен використати команду `pip install /path/to/wheelfile.whl`. Результат встановлення бібліотеки предствлено на рисунку 3.16. Після виконання згаданих команд бібліотека стане доступною для використання в клієнтському застосунку. Для цього користувачу потрібно додати у код свого програмного продукту рядок `from original_least_squares_boost.original_least_squares_boost import MultivariableLinearRegressionModel`.

Для використання системи статистичного імітаційного моделювання ефективності досліджуваного методу користувачу спочатку доведеться розгорнути роброблену бібліотеку. Далі користувач може клонувати репозиторій з кодом застосунку. Наступним кроком може бути запуск програмного за допомогою команди `python .\effectiveness_research_system.py`. На рисунку 3.17 представлено головне вікно застосунку, яке побачить користувач після запуску.

```

PS D:\marictp\src\library\original_least_squares_boost> python setup.py bdist_wheel
running bdist_wheel
running build
running build_py
copying original_least_squares_boost\original_least_squares_boost.py -> build\lib\original_least_squares_boost
C:\Users\Volodymyr_Hrytsiuk\AppData\Local\Programs\Python\Python311\Lib\site-packages\setuptools\command\install.py:34: Setuptools-based tools.
  warnings.warn(
installing to build\bdist.win-amd64\wheel
running install
running install_lib
creating build\bdist.win-amd64\wheel
creating build\bdist.win-amd64\wheel\original_least_squares_boost
copying build\lib\original_least_squares_boost\original_least_squares_boost.py -> build\bdist.win-amd64\wheel\.\original_le
copying build\lib\original_least_squares_boost\__init__.py -> build\bdist.win-amd64\wheel\.\original_least_squares_boost
running install_egg_info
running egg_info
writing original_least_squares_boost.egg-info\PKG-INFO
writing dependency links to original_least_squares_boost.egg-info\dependency_links.txt
writing top-level names to original_least_squares_boost.egg-info\top_level.txt
reading manifest file 'original_least_squares_boost.egg-info\SOURCES.txt'
writing manifest file 'original_least_squares_boost.egg-info\SOURCES.txt'
Copying original_least_squares_boost.egg-info to build\bdist.win-amd64\wheel\.\original_least_squares_boost-1.12-py3.11.egg
running install_scripts
creating build\bdist.win-amd64\wheel\original_least_squares_boost-1.12.dist-info\WHEEL
creating 'dist\original_least_squares_boost-1.12-py3-none-any.whl' and adding 'build\bdist.win-amd64\wheel' to it
adding 'original_least_squares_boost/__init__.py'
adding 'original_least_squares_boost/original_least_squares_boost.py'
adding 'original_least_squares_boost-1.12.dist-info\METADATA'
adding 'original_least_squares_boost-1.12.dist-info\WHEEL'
adding 'original_least_squares_boost-1.12.dist-info\top_level.txt'
adding 'original_least_squares_boost-1.12.dist-info\RECORD'
removing build\bdist.win-amd64\wheel

```

Рисунок 3.15 – Збірка бібліотеки

```

PS D:\marictp\src\library\original_least_squares_boost> pip install .\dist\original_least_squares_boost-1.15-py3-none-any.whl
Processing d:\marictp\src\library\original_least_squares_boost\dist\original_least_squares_boost-1.15-py3-none-any.whl
Installing collected packages: original-least-squares-boost
  Attempting uninstall: original-least-squares-boost
    Found existing installation: original-least-squares-boost 1.12
    Uninstalling original-least-squares-boost-1.12:
      Successfully uninstalled original-least-squares-boost-1.12
  Successfully installed original-least-squares-boost-1.15
Successfully installed original-least-squares-boost-1.15

```

Рисунок 3.16 – Встановлення бібліотеки

Application

Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data

Number of independent values (m, 1..30):

Number of zero coefficients (1..m):

Number of experiments:

Number of repetitions of the experiment (10..1000):

E-distribution: Uniform Normal Uniform + Normal

Uniform distributions parameter (1..10):

MLR coefficients:

b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
0	0	0	0	0	0	0	0	0	0	0

Рисунок 3.17 – Початковий стан застосунку

Якщо користувач бажає отримати виконуваний файл (exe), то повинен використати команду `«pyinstaller .\effectivness_research_system.py»`. Результат роботи застосунку не залежить від способу, за допомогою якого він запускається.

Висновки до розділу

В третьому розділі описано програмне забезпечення, що включає в себе бібліотеку, що реалізує оригінальний метод побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних, та застосунок, що є системою статистичного імітаційного моделювання ефективності досліджуваного методу. Для цього було обґрунтовано, що монолітна архітектура є найкращим варіантом для обох частин програмного забезпечення. Також було представлено діаграми класів обох програмних продуктів та описано ключові зони відповідальності кожної сутності. В межах даного розділу також було описано та проілюстровано функціональність розробленого програмного забезпечення. Для користувача надано зручну інструкцію щодо розгортання бібліотеки та системи статистичного імітаційного моделювання ефективності досліджуваного методу.

4 МОДЕЛЮВАННЯ ЕФЕКТИВНОСТІ ЗАСТОСУВАННЯ ДОСЛІДЖУВНОГО МЕТОДА ЗА ДОПОМОГОЮ РОЗРОБЛЕНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

4.1 Процедура проведення досліджень

В першу чергу потрібно дослідити ефективність застосування методу при зміні значень випадкової величини E та зміні кількості вхідних змінних. Основну увагу при дослідженні потрібно приділяти метрикам ефективності оцінок коефіцієнтів багатовимірної лінійної регресії, а саме:

- відсоток правильно знайдених структур БЛР;
- значення інтегральної похибки компонент вектора оцінок, знайдених за допомогою загальної процедури МНК, відносно фактичних значень, заданих при моделюванні;
- значення інтегральної похибки компонент вектора оцінок, знайдених за допомогою досліджуваного методу, відносно фактичних значень, заданих при моделюванні;
- відсоток покращення ефективності оцінки коефіцієнтів досліджуваним методом відносно загальної процедури метода найменших квадратів.

Дослідження буде проводитись для багатовимірних лінійних регресійних моделей розмірністю 10, 15, 20. Кількість нульових коефіцієнтів у кожному випадку буде задаватися як 40%, 60% та 80% від кількості незалежних вхідних змінних БЛР.

Кількість випробувань в імітації одного активного експерименту задається як x_2 , x_3 , x_4 , x_5 від кількості незалежних вхідних змінних БЛР.

Задається три види розподілу випадкової величини E : рівномірний на відрізьку від $-c$ до c , нормальний, сума нормального та рівномірного розподілу. Для рівномірного розподілу, значення параметра c задаються числами від 1 до 10. Для нормального розподілу математичне очікування задається 0, а стандартне відхилення (сігма) задається значеннями від 1 до 10.

Фактичні значення коефіцієнтів БЛР при імітації генеруються випадковим чином в інтервалі від -2 до -0,5 та від 0,5 до 2.

Значення незалежних змінних при імітації генеруються випадковим чином в інтервалі від 1 до 20.

Кількість повторів експерименту при фіксованих параметрах моделювання – 1000.

За результатами моделювання особлива увага приділяється відсотку правильно знайдених структур БЛР. Для побудови рекомендацій з застосування досліджуваного методу будуть фійсуватися параметри розмірності БЛР, кількості випробовувань та розподілу випадкової величини E , при яких відсоток правильно знайдених моделей складає більше 70%.

4.2 Результати статистичного моделювання ефективності застосування досліджуваного методу побудови БЛР при рівномірному розподілі випадкової величини E

За допомогою розробленого програмного забезпечення отримано результати статистичного моделювання ефективності застосування досліджуваного методу побудови БЛР, заданої надлишковим описом, на малому об'ємі експериментальних даних, при рівномірному розподілі випадкової величини E , що відображені в таблицях 4.1 – 4.36. У першій колонці зазначено параметр рівномірного розподілу у відповідності до того, як він задається у розробленому застосунку для статистичного імітаційного моделювання досліджуваного методу. На основі цього параметру визначається інтервал рівномірного розподілу. Наприклад, якщо задано $c=2$, то значення випадкової величини E будуть згенеровані відповідно до рівномірного розподілу в інтервалі від -2 до 2. У другій колонці вказано інтегральну похибку вектора оцінок, знайдених за допомогою загальної процедури МНК. У третій колонці вказано інтегральну похибку вектора оцінок, знайдених за допомогою досліджуваного методу. Четверта колонка містить відсоток покращення ефективності оцінок коефіцієнтів багатовимірної лінійної регресії за допомогою оригінального методу відносно загальної процедури МНК.

Таблиця 4.1 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінні не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок
1	100.0	0.22	0.14	36.1
2	99.5	0.38	0.28	28.3
3	95.3	0.52	0.38	27.6
4	87.9	0.61	0.46	25.0
5	78.5	0.69	0.51	26.2
6	70.4	0.73	0.58	20.5
7	59.3	0.79	0.62	21.8

Таблиця 4.2 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінні не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100.0	0.16	0.11	29.8
2	100.0	0.29	0.22	25.9
3	99.0	0.42	0.30	27.4
4	96.5	0.50	0.37	24.9
5	87.5	0.57	0.46	20.3

Продовження таблиці 4.2

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
6	82.2	0.64	0.50	21.0
7	77.6	0.70	0.52	25.7
8	68.8	0.68	0.56	18.0
9	59.8	0.76	0.61	18.7
10	54.5	0.77	0.63	18.3

Таблиця 4.3 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінні не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100.0	0.13	0.09	27.0
2	100.0	0.24	0.18	27.2
3	99.8	0.34	0.26	25.7
4	97.9	0.44	0.33	24.3
5	94.0	0.51	0.40	20.7
6	89.6	0.55	0.44	19.0
7	83.8	0.60	0.48	20.4
8	79.8	0.64	0.50	20.7
9	72.7	0.69	0.55	20.3
10	68.3	0.73	0.59	18.9

Таблиця 4.4 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінні не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100.0	0.11	0.08	24.4
2	100.0	0.22	0.16	26.7
3	100.0	0.30	0.23	22.7
4	99.2	0.38	0.30	21.8
5	97.2	0.46	0.36	21.8
6	93.6	0.50	0.41	18.8
7	89.3	0.57	0.45	20.5
8	83.4	0.59	0.47	20.0
9	78.6	0.65	0.52	19.0
10	75.4	0.67	0.55	19.1

Таблиця 4.5 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.9	0.27	0.13	50.4
2	96.1	0.46	0.25	46.1

Продовження таблиці 4.5

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
3	85.5	0.57	0.33	42.2
4	75.1	0.70	0.40	42.9
5	62.9	0.74	0.44	40.4
6	54.9	0.73	0.50	31.9

Таблиця 4.6 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.18	0.10	47.2
2	99.6	0.35	0.20	41.7
3	94.4	0.49	0.29	40.7
4	86.4	0.57	0.33	41.2
5	79	0.64	0.38	40.4
6	70.5	0.69	0.43	37.4
7	61.8	0.71	0.49	31.0
8	54.4	0.74	0.49	33.9
9	52.4	0.81	0.53	34.1

Таблиця 4.7 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.15	0.08	42.8
2	100	0.28	0.17	40.5
3	97.3	0.41	0.25	39.0
4	90.6	0.48	0.28	40.4
5	84.4	0.56	0.36	36.6
6	79	0.60	0.37	37.5
7	71.6	0.67	0.41	38.6
8	64.3	0.67	0.47	30.8
9	66.1	0.75	0.49	34.6
10	51.7	0.76	0.51	32.7

Таблиця 4.8 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.13	0.08	38.5
2	100	0.25	0.15	39.8

Продовження таблиці 4.8

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
3	99.2	0.35	0.22	37.1
4	94.4	0.44	0.27	37.9
5	89	0.49	0.32	35.8
6	83	0.57	0.36	36.4
7	78.2	0.63	0.41	35.1
8	74.1	0.66	0.44	34.2
9	63.8	0.68	0.46	32.6
10	65.8	0.71	0.48	32.3

Таблиця 4.9 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.6	0.31	0.08	73.1
2	99.9	0.54	0.19	64.7
3	99.3	0.66	0.28	57.8
4	94.1	0.79	0.39	50.7
5	82.1	0.90	0.51	42.9
6	65.3	0.92	0.62	32.4

Таблиця 4.10 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.6	0.19	0.07	64.1
2	99.9	0.42	0.14	65.4
3	100	0.57	0.22	61.2
4	99.3	0.64	0.30	53.8
5	94.5	0.75	0.40	47.6
6	85.4	0.80	0.45	42.9
7	77.3	0.86	0.56	35.2
8	62	0.87	0.61	29.9
9	54.4	0.90	0.66	27.1

Таблиця 4.11 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.4	0.15	0.06	63.5
2	99.9	0.32	0.12	62.5
3	99.9	0.48	0.18	62.1

Продовження таблиці 4.11

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
4	99.8	0.59	0.29	51.6
5	97.8	0.64	0.33	49.2
6	94.9	0.74	0.38	49.5
7	88	0.78	0.46	41.0
8	78.6	0.83	0.51	38.5
9	64.8	0.84	0.57	32.4
10	56.2	0.83	0.60	27.8

Таблиця 4.12 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.9	0.13	0.05	60.5
2	100	0.27	0.10	61.9
3	100	0.42	0.17	60.3
4	100	0.54	0.24	55.4
5	99.8	0.67	0.30	55.7
6	97.2	0.71	0.37	48.1
7	92.9	0.74	0.40	45.4
8	85.5	0.80	0.47	41.7

Продовження таблиці 4.12

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
9	80	0.86	0.54	36.6
10	68.9	0.89	0.60	32.9

Таблиця 4.13 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 45 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.16	0.11	30.3
2	98.8	0.30	0.22	26.0
3	91.2	0.42	0.31	25.2
4	84.7	0.53	0.41	21.4
5	75.4	0.60	0.48	20.6
6	67.6	0.67	0.54	19.1
7	58.2	0.77	0.59	22.3
8	53.8	0.78	0.65	16.0

Таблиця 4.14 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.8	0.22	0.15	31.7
2	95	0.41	0.27	33.5
3	82.5	0.55	0.38	29.6
4	70.6	0.64	0.48	23.8
5	59.9	0.73	0.59	19.5

Таблиця 4.15 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.13	0.09	27.3
2	99.4	0.26	0.19	26.7
3	97	0.35	0.26	25.5
4	89	0.46	0.34	24.9
5	80.2	0.54	0.42	23.0
6	75.6	0.60	0.46	24.3
7	66.3	0.64	0.51	20.2

Продовження таблиці 4.15

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
8	62.4	0.70	0.56	20.1
9	61.2	0.72	0.61	16.2
10	54.3	0.79	0.65	17.2

Таблиця 4.16 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 75 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.11	0.08	26.1
2	99.7	0.22	0.17	24.6
3	97.9	0.32	0.24	23.7
4	93.6	0.41	0.31	24.2
5	88.6	0.47	0.38	19.3
6	81.8	0.54	0.43	20.2
7	76.3	0.60	0.49	18.9
8	71.3	0.66	0.54	17.7
9	64.5	0.69	0.58	16.4
10	60.4	0.73	0.62	15.9

Таблиця 4.17 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.20	0.11	46.9
2	98.4	0.38	0.20	46.6
3	93.3	0.53	0.30	43.8
4	88	0.61	0.39	36.2
5	80.2	0.70	0.47	32.9
6	76	0.79	0.52	35.0
7	71	0.84	0.63	25.6
8	67.8	0.93	0.68	27.4
9	63	0.94	0.70	24.8
10	50.7	0.99	0.76	23.2

Таблиця 4.18 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 45 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.15	0.08	46.0
2	99.9	0.27	0.16	43.1

Продовження таблиці 4.18

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
3	96.9	0.40	0.23	40.7
4	92	0.51	0.32	37.7
5	85.6	0.60	0.40	34.4
6	83.8	0.69	0.47	31.7
7	81.8	0.73	0.52	28.3
8	78.1	0.76	0.54	29.1
9	73	0.79	0.58	26.5
10	69.3	0.89	0.66	26.4

Таблиця 4.19 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.12	0.07	42.7
2	100	0.22	0.14	36.7
3	99.2	0.33	0.20	38.3
4	95.5	0.42	0.26	38.2
5	92.7	0.50	0.32	35.1
6	89.5	0.58	0.39	32.7
7	83.4	0.64	0.45	30.4

Продовження таблиці 4.19

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
8	79.9	0.69	0.50	27.5
9	79.4	0.74	0.54	26.4
10	73.2	0.77	0.58	24.2

Таблиця 4.20 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 75 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.10	0.06	39.8
2	100	0.20	0.12	39.1
3	99.5	0.31	0.19	39.7
4	97.7	0.38	0.24	35.7
5	94.2	0.46	0.29	37.7
6	91.2	0.50	0.34	32.1
7	85.4	0.58	0.40	30.6
8	86.1	0.64	0.44	30.4
9	82.8	0.67	0.49	26.5
10	77.5	0.75	0.54	27.9

Таблиця 4.21 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.34	0.12	64.9
2	97.6	0.60	0.23	61.1
3	89.9	0.74	0.34	53.8
4	83.2	0.81	0.43	47.1
5	77.8	0.89	0.53	40.2
6	71.6	1.00	0.58	41.8
7	67.4	1.01	0.65	35.6
8	60.1	1.05	0.73	30.3
9	52.2	1.08	0.78	28.2

Таблиця 4.22 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 45 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.24	0.10	60.1
2	99.5	0.42	0.20	53.5
3	94.5	0.58	0.27	52.7

Продовження таблиці 4.22

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
4	89.8	0.69	0.36	47.3
5	82.1	0.78	0.41	47.3
6	80.4	0.85	0.50	41.4
7	75.1	0.87	0.57	35.2
8	70.7	0.97	0.63	35.5
9	67.5	0.97	0.68	29.8
10	60.5	0.97	0.70	27.8

Таблиця 4.23 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.20	0.08	58.4
2	99.9	0.37	0.17	54.0
3	97.3	0.54	0.25	53.5
4	93.4	0.62	0.30	50.8
5	87	0.68	0.38	44.2
6	81.8	0.78	0.45	42.2
7	79.7	0.85	0.49	42.4
8	76.4	0.87	0.56	35.9

Продовження таблиці 4.23

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
9	69.9	0.92	0.61	33.2
10	67.5	0.95	0.64	32.1

Таблиця 4.24 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 75 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.17	0.08	55.4
2	99.9	0.32	0.14	55.1
3	98.5	0.46	0.21	54.5
4	94.4	0.56	0.28	50.5
5	89.4	0.66	0.36	45.0
6	86	0.72	0.41	43.3
7	82.7	0.81	0.45	44.3
8	78.5	0.82	0.52	36.2
9	75.9	0.86	0.54	36.4
10	72.2	0.90	0.61	32.4

Таблиця 4.25 статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	92.3	0.18	0.12	32.9
2	74.1	0.34	0.23	31.9
3	65.5	0.47	0.32	33.2
4	59.4	0.54	0.40	26.8
5	52.4	0.62	0.47	23.5

Таблиця 4.26 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	97.1	0.13	0.09	28.4
2	81.4	0.25	0.18	30.3
3	69.7	0.34	0.26	24.6
4	66.4	0.46	0.32	30.3
5	59.6	0.51	0.40	21.9
6	55.5	0.58	0.44	24.9

Таблиця 4.27 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 80 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	98.2	0.11	0.08	27.5
2	84.7	0.22	0.16	25.5
3	76.7	0.30	0.23	22.5
4	68.8	0.37	0.28	25.9
5	62.2	0.44	0.33	24.7
6	58.7	0.50	0.38	24.7
7	57.5	0.54	0.43	20.9
8	54.3	0.58	0.45	21.8
9	45.8	0.63	0.50	20.8
10	42.1	0.72	0.53	25.6

Таблиця 4.28 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 100 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.4	0.09	0.07	28.3
2	89.2	0.18	0.13	24.2

Продовження таблиці 4.28

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
3	78.2	0.27	0.20	27.5
4	70.8	0.33	0.25	24.7
5	68.3	0.39	0.30	23.4
6	61.8	0.46	0.35	22.5
7	60.8	0.50	0.38	22.9
8	57.2	0.55	0.44	20.7
9	50.8	0.60	0.46	23.1
10	50.5	0.61	0.51	17.0

Таблиця 4.29 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.9	0.24	0.12	51.2
2	95.8	0.42	0.22	46.4
3	87.9	0.54	0.33	39.7
4	81.1	0.66	0.42	37.2
5	75	0.74	0.47	36.9
6	70.1	0.80	0.53	33.4
7	65.8	0.83	0.58	30.1

Продовження таблиці 4.29

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
8	61.3	0.90	0.63	29.8
9	57.1	0.91	0.68	25.1
10	52	0.94	0.70	24.9

Таблиця 4.30 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.16	0.09	41.3
2	98.8	0.30	0.18	39.0
3	94.7	0.43	0.26	39.6
4	86.8	0.52	0.32	38.2
5	82.5	0.61	0.40	34.0
6	77.1	0.67	0.45	33.0
7	71.2	0.74	0.51	30.8
8	70.3	0.78	0.58	25.7
9	67.2	0.82	0.60	27.0
10	61	0.86	0.66	23.0

Таблиця 4.31 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 80 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.14	0.08	41.6
2	99.4	0.25	0.16	37.0
3	97	0.35	0.22	37.6
4	90.8	0.45	0.30	34.1
5	86.6	0.52	0.35	33.4
6	80.7	0.59	0.41	30.7
7	75	0.63	0.44	30.5
8	74.8	0.71	0.49	30.2
9	72.1	0.74	0.55	25.0
10	67.8	0.79	0.58	26.3

Таблиця 4.32 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 100 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.11	0.07	37.7
2	99.7	0.23	0.13	41.1

Продовження таблиці 4.32

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
3	97.3	0.33	0.20	39.9
4	94	0.40	0.24	40.1
5	88.3	0.47	0.31	33.6
6	85	0.54	0.36	32.6
7	80	0.58	0.41	29.1
8	79.1	0.63	0.44	29.7
9	74.5	0.67	0.51	25.0
10	68.5	0.75	0.53	29.5

Таблиця 4.33 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 16 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.27	0.10	64.0
2	100	0.48	0.19	59.2
3	100	0.62	0.27	56.8
4	99.8	0.72	0.36	50.5
5	99.8	0.80	0.41	48.1
6	98.2	0.85	0.47	45.2
7	96.8	0.92	0.51	45.3

Продовження таблиці 4.33

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
8	94.7	0.96	0.58	39.6
9	92	0.98	0.59	39.4
10	89	1.02	0.63	38.4

Таблиця 4.34 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 16 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.20	0.08	60.6
2	100	0.36	0.15	57.3
3	100	0.49	0.23	53.0
4	100	0.60	0.29	51.9
5	100	0.66	0.34	49.3
6	99.7	0.74	0.40	45.7
7	99.1	0.76	0.44	42.4
8	97.9	0.85	0.48	43.3
9	96.5	0.86	0.52	39.9
10	94.3	0.89	0.56	37.8

Таблиця 4.35 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 16 змінних не впливають на значення вихідної змінної при 80 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.16	0.07	58.7
2	100	0.30	0.13	56.3
3	100	0.42	0.19	55.6
4	100	0.52	0.24	52.9
5	100	0.59	0.30	49.3
6	100	0.64	0.37	42.8
7	99.7	0.73	0.40	44.8
8	99	0.77	0.42	45.0
9	97.7	0.80	0.49	38.6
10	97	0.86	0.50	41.9

Таблиця 4.36 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 20 незалежних вхідних змінних, з яких 16 змінних не впливають на значення вихідної змінної при 100 випробуваннях в одному експерименті

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.14	0.06	57.9
2	100	0.26	0.12	54.9

Продовження таблиці 4.36

с	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
3	100	0.37	0.18	52.1
4	100	0.48	0.23	52.8
5	100	0.52	0.27	49.3
6	99.9	0.60	0.33	45.0
7	100	0.64	0.37	42.3
8	99.2	0.71	0.40	43.7
9	99.6	0.74	0.43	42.2
10	98.6	0.80	0.47	40.7

4.3 Результати статистичного моделювання ефективності застосування досліджуваного метода побудови БЛР при нормальному розподілі випадкової величини Е

У таблицях 4.37 – 4.60 представлено результати моделювання ефективності метода при нормальному розподілі Е. У першій колонці зазначено стандартне відхилення (σ) нормального розподілу.

Таблиця 4.37 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	99.8	0.39	0.27	31.4

Продовження таблиці 4.37

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
2	96.6	0.63	0.47	24.8
3	81.3	0.72	0.56	21.9
4	61.7	0.80	0.65	19.0

Таблиця 4.38 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.28	0.20	29.6
2	98.9	0.50	0.39	21.6
3	95	0.64	0.52	19.1
4	86.7	0.71	0.58	17.9
5	72.1	0.78	0.64	17.8
6	54.2	0.81	0.72	11.4
7	40	0.91	0.74	18.3
8	32.3	0.94	0.84	10.7
9	22.2	0.94	0.91	2.3
10	14.1	0.94	0.88	6.6

Таблиця 4.39 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.23	0.17	26.5
2	99.9	0.43	0.32	26.1
3	98.1	0.58	0.45	21.9
4	91.8	0.67	0.54	18.5
5	83.3	0.72	0.60	16.7
6	71.3	0.77	0.66	14.7
7	63.7	0.83	0.70	16.3

Таблиця 4.40 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.20	0.15	29.0
2	100	0.39	0.29	25.3
3	99.1	0.52	0.42	20.6
4	94	0.63	0.51	18.6
5	87.6	0.68	0.55	18.7

Продовження таблиці 4.40

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
6	80.1	0.75	0.64	14.7
7	69.2	0.80	0.66	16.7
8	60	0.80	0.71	10.4
9	49.8	0.82	0.75	8.6
10	41.2	0.88	0.79	9.8

Таблиця 4.41 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.39	0.20	48.5
2	100	0.61	0.38	37.6
3	99.6	0.76	0.50	34.3
4	98.5	0.84	0.61	27.0
5	90.2	0.94	0.68	27.6
6	76.6	0.98	0.74	24.2
7	63.9	0.96	0.77	20.0
8	49.3	1.00	0.82	17.7

Таблиця 4.42 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.28	0.16	41.4
2	100	0.50	0.31	38.0
3	100	0.66	0.45	31.7
4	99.8	0.76	0.52	31.2
5	98.7	0.82	0.62	24.4
6	95.9	0.88	0.70	20.6
7	87.9	0.91	0.74	19.6
8	79.5	0.95	0.75	21.4
9	68	0.98	0.79	19.4
10	55.7	0.99	0.80	19.8

Таблиця 4.43 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.23	0.14	40.0
2	100	0.43	0.27	37.4

Продовження таблиці 4.43

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
3	100	0.56	0.38	33.3
4	100	0.68	0.49	28.0
5	100	0.77	0.56	26.9
6	99.2	0.82	0.62	23.5
7	97.3	0.86	0.70	18.7
8	91.2	0.91	0.69	24.6
9	85.4	0.93	0.77	17.6
10	76.1	0.94	0.79	16.3

Таблиця 4.44 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.20	0.12	40.0
2	100	0.38	0.24	38.6
3	100	0.52	0.35	31.2
4	100	0.61	0.42	30.7
5	100	0.70	0.51	27.3
6	99.5	0.74	0.55	25.5
7	98.8	0.82	0.63	23.4

Продовження таблиці 4.44

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
8	96.6	0.84	0.66	21.7
9	93.1	0.91	0.72	20.8
10	89.6	0.92	0.75	18.0

Таблиця 4.45 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.48	0.18	63.2
2	100	0.77	0.36	53.1
3	96.8	0.89	0.49	45.0
4	82.9	0.94	0.59	36.8
5	61.7	0.96	0.67	30.5
6	43.1	1.00	0.70	30.7
7	29.8	1.09	0.87	20.5
8	20.3	1.11	0.89	19.7
9	13.5	1.13	0.87	23.0

Таблиця 4.46 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.37	0.14	63.1
2	100	0.63	0.28	55.2
3	99.6	0.78	0.42	46.4
4	96.7	0.89	0.51	42.5
5	85.9	0.92	0.61	33.8
6	72.4	0.94	0.64	32.1
7	56.5	0.98	0.73	25.9

Таблиця 4.47 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.30	0.12	58.9
2	100	0.51	0.24	51.7
3	100	0.69	0.34	50.2
4	98.9	0.77	0.47	39.3
5	94.7	0.86	0.51	40.8

Продовження таблиці 4.47

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
7	73	0.93	0.64	31.2
8	57.2	0.99	0.76	23.7

Таблиця 4.48 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.25	0.11	56.8
2	100	0.47	0.22	52.8
3	100	0.61	0.32	48.1
4	99.7	0.73	0.41	44.1
5	98	0.79	0.49	38.5
6	93.9	0.84	0.56	33.8
7	82.8	0.94	0.64	31.6
8	72.4	0.91	0.68	25.4
9	60.6	0.98	0.72	26.4
10	49.7	0.97	0.74	23.8

Таблиця 4.49 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.32	0.21	33.4
2	99.5	0.52	0.39	26.2
3	95	0.66	0.51	22.4
4	81.8	0.75	0.58	22.8
5	66.8	0.81	0.66	19.3
6	54.8	0.83	0.69	17.2

Таблиця 4.50 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 45 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.24	0.17	29.7
2	100	0.41	0.32	21.4
3	98.9	0.54	0.43	21.2
4	95.1	0.64	0.51	19.7
5	89.6	0.71	0.56	20.9
6	78.5	0.76	0.62	18.5

Таблиця 4.51 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.19	0.13	27.6
2	100	0.35	0.27	23.4
3	100	0.48	0.37	23.5
4	97.8	0.56	0.48	14.9
5	96.3	0.65	0.53	18.4
6	90.7	0.70	0.59	15.7
7	81.7	0.74	0.61	17.3
8	72	0.79	0.67	15.2
9	61.3	0.82	0.67	18.1

Таблиця 4.52 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 75 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.16	0.12	27.3
2	100	0.31	0.24	23.4
3	100	0.44	0.34	23.0

Продовження таблиці 4.52

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
4	99.5	0.52	0.43	17.0
5	98.1	0.61	0.49	18.4
6	92.9	0.66	0.55	16.7
7	87.9	0.69	0.58	15.4
8	83.9	0.77	0.62	19.7
9	75	0.78	0.65	15.8
10	65.9	0.77	0.68	11.5

Таблиця 4.53 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	96	0.36	0.18	48.6
2	75	0.62	0.38	39.7
3	52.1	0.82	0.54	34.5
4	38.2	0.96	0.77	19.9

Таблиця 4.54 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 45 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	98.8	0.27	0.15	43.8
2	85.1	0.48	0.30	36.4
3	66.7	0.63	0.45	28.6
4	50.4	0.78	0.58	24.8

Таблиця 4.55 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.22	0.13	42.6
2	91.6	0.42	0.26	39.3
3	75.2	0.56	0.38	32.7
4	59.1	0.70	0.52	25.8
5	49.7	0.80	0.60	25.3
6	40.7	0.91	0.74	19.2

Таблиця 4.56 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 75 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.18	0.11	38.3
2	92	0.35	0.22	36.1
3	80.1	0.49	0.33	33.7
4	68.9	0.66	0.46	30.0
5	55.3	0.73	0.55	24.5
6	45.2	0.82	0.62	24.0

Таблиця 4.57 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.42	0.16	63.1
2	100	0.68	0.30	55.4
3	98.8	0.80	0.41	49.0
4	91.7	0.90	0.48	46.3
5	79.3	0.91	0.54	41.1
6	65.2	0.97	0.59	38.4

Таблиця 4.58 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 45 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.32	0.13	60.9
2	100	0.54	0.25	53.2
3	99.4	0.67	0.36	46.5
4	98.5	0.78	0.42	45.5
5	93.9	0.84	0.47	43.7
6	86.3	0.90	0.52	41.8
7	76.2	0.95	0.58	39.4
8	65	0.96	0.60	36.8

Таблиця 4.59 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.26	0.11	58.1
2	100	0.48	0.22	54.3
3	99.9	0.62	0.31	49.9
4	99.7	0.73	0.40	44.5

Продовження таблиці 4.59

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
5	97.2	0.79	0.47	40.5
6	91.8	0.83	0.50	39.5
7	88	0.87	0.54	37.8
8	81.1	0.90	0.55	38.9
9	69.2	0.94	0.60	35.7

Таблиця 4.60 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 12 змінних не впливають на значення вихідної змінної при 75 випробуваннях в одному експерименті

σ	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	100	0.23	0.10	56.9
2	100	0.43	0.19	56.4
3	100	0.57	0.28	51.8
4	99.8	0.67	0.37	44.9
5	99.2	0.76	0.43	43.7
6	97.2	0.81	0.48	40.4
7	92.6	0.85	0.52	39.4
8	87.2	0.88	0.52	40.8
9	81.1	0.87	0.57	33.8

4.4 Результати статистичного моделювання ефективності застосування досліджуваного метода побудови БЛР при сумі нормального і рівномірного розподілу величини E

У таблицях 4.61-4.80 представлено результати статистичного імітаційного моделювання ефективності досліджуваного метода побудови БЛР при сумі нормального і рівномірного розподілу випадкової величини E .

Таблиця 4.61 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	93.7	0.36	0.27	26.5
2	2	77.6	0.59	0.44	25.2
3	3	64.8	0.70	0.56	20.5

Таблиця 4.62 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 45 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.1	0.27	0.20	26.3

Продовження таблиці 4.62

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
2	2	85.7	0.46	0.35	23.9
3	3	72.6	0.58	0.46	20.4
4	4	60	0.72	0.57	20.7
5	5	56.3	0.76	0.62	18.6

Таблиця 4.63 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.5	0.23	0.17	26.6
2	2	90.2	0.40	0.30	23.3
3	3	79.1	0.50	0.41	19.1
4	4	69.1	0.63	0.50	20.3
5	5	63.7	0.71	0.56	21.2
6	6	54	0.73	0.63	14.4
7	7	51.3	0.80	0.68	14.5

Таблиця 4.64 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 75 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	100	0.20	0.15	23.2
2	2	91.3	0.36	0.28	22.0
3	3	82.9	0.48	0.38	20.9
4	4	73.7	0.58	0.47	19.4
5	5	65.3	0.67	0.55	18.6
6	6	59.5	0.67	0.56	16.2
7	7	53.3	0.78	0.66	15.6

Таблиця 4.65 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	88.3	0.38	0.20	47.0
2	2	70.6	0.59	0.36	39.6
3	3	60.2	0.71	0.48	32.7

Таблиця 4.66 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 45 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	93.5	0.28	0.16	42.6
2	2	78	0.48	0.30	38.4
3	3	67.9	0.61	0.40	35.4
4	4	56.7	0.71	0.47	34.0
5	5	53.9	0.77	0.56	27.4

Таблиця 4.67 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 60 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	96.7	0.24	0.14	42.3
2	2	81.6	0.40	0.25	36.4
3	3	66.9	0.54	0.34	37.4
4	4	63.3	0.63	0.44	29.9
5	5	60.8	0.71	0.50	29.6

Таблиця 4.68 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 15 незалежних вхідних змінних, з яких 9 змінних не впливають на значення вихідної змінної при 75 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	97.5	0.21	0.12	42.4
2	2	84.6	0.37	0.23	37.4
3	3	75.9	0.50	0.32	35.3
4	4	68.2	0.60	0.40	33.8
5	5	64.7	0.71	0.47	33.4
6	6	60.5	0.72	0.53	27.1
7	7	59	0.78	0.56	28.3

Таблиця 4.69 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	97.7	0.55	0.30	44.3
2	2	84.7	0.81	0.55	32.3
3	3	65.2	0.99	0.78	21.2

Таблиця 4.70 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.6	0.42	0.24	42.5
2	2	90.2	0.71	0.44	37.5
3	3	79.6	0.84	0.63	24.4
4	4	68.4	0.96	0.75	22.2
5	5	51.6	1.04	0.87	16.6

Таблиця 4.71 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.9	0.34	0.20	40.0
2	2	93.5	0.62	0.41	34.8
3	3	87	0.75	0.56	25.4
4	4	76.2	0.91	0.68	24.4
5	5	62.5	1.00	0.80	20.1

Таблиця 4.72 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 6 змінних не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	100	0.30	0.19	37.2
2	2	96.2	0.54	0.37	32.3
3	3	88.8	0.71	0.50	29.6
4	4	80	0.84	0.64	24.1
5	5	72.7	0.88	0.69	22.1
6	6	60.8	0.97	0.83	15.1
7	7	50.8	1.00	0.83	16.4

Таблиця 4.73 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	100	0.53	0.22	57.5
2	2	99.8	0.77	0.39	49.2
3	3	96.3	0.92	0.53	42.5

Продовження таблиці 4.73

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
4	4	83.3	0.94	0.56	40.4
5	5	59.8	1.01	0.65	35.7

Таблиця 4.74 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.9	0.39	0.17	56.3
2	2	99.9	0.64	0.32	49.6
3	3	99.8	0.78	0.47	39.5
4	4	96.1	0.88	0.54	39.4
5	5	84.9	0.94	0.61	35.0
6	6	67	0.97	0.66	31.5
7	7	51.8	0.99	0.67	33.0

Таблиця 4.75 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.8	0.34	0.15	57.0
2	2	99.9	0.56	0.28	49.8
3	3	99.8	0.73	0.40	45.4
4	4	99.2	0.82	0.49	40.4
5	5	94	0.87	0.55	37.3
6	6	85.1	0.94	0.63	32.7
7	7	67.3	0.94	0.63	33.0
8	8	53.4	0.97	0.67	31.5

Таблиця 4.76 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 8 змінних не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.9	0.30	0.13	56.3
2	2	100	0.52	0.26	50.6

Продовження таблиці 4.76

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
3	3	100	0.65	0.36	43.9
4	4	99.8	0.75	0.45	39.1
5	5	98.5	0.87	0.54	37.5
6	6	91.2	0.92	0.58	36.2
7	7	82.7	0.92	0.61	34.3
8	8	67.7	0.96	0.65	31.8
9	9	55.7	0.99	0.69	30.2

Таблиця 4.77 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 20 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	93	0.42	0.29	30.8
2	2	65.4	0.65	0.50	23.5
3	3	43.8	0.78	0.62	20.5

Таблиця 4.78 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 30 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.1	0.33	0.25	25.2
2	2	81.7	0.54	0.42	22.2
3	3	62	0.64	0.54	15.7
4	4	47.5	0.74	0.61	17.9

Таблиця 4.79 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 40 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.6	0.26	0.20	24.1
2	2	89.8	0.46	0.37	18.3
3	3	73.4	0.60	0.47	21.4
4	4	58.7	0.66	0.55	16.9
5	5	47.1	0.74	0.64	13.2

Таблиця 4.80 – Результати статистичного моделювання ефективності застосування досліджуваного метода для побудови БЛР, що містить 10 незалежних вхідних змінних, з яких 4 змінних не впливають на значення вихідної змінної при 50 випробуваннях в одному експерименті

σ	c	Відсоток правильно знайдених структур БЛР	Інтегральна похибка (МНК)	Інтегральна похибка (досліджуваний метод)	Відсоток покращення ефективності оцінок коефіцієнтів БЛР
1	1	99.9	0.24	0.18	24.5
2	2	93	0.42	0.34	19.7
3	3	81.7	0.54	0.45	16.7
4	4	67.8	0.65	0.55	16.0
5	5	53.4	0.70	0.59	15.5

Висновки до розділу

В даному розділі розглянуто і проаналізовано результати статистичного моделювання ефективності досліджуваного методу побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних. У розділі наведено таблиці, що дозволяють зробити висновок про область ефективного застосування досліджуваного метода. Таким чином, варто відмітити, що збільшення кількості коефіцієнтів призводить до зменшення відсотку правильно знайдених структур БЛР. Окрім того, і збільшення кількості нульових коефіцієнтів також призводить до невеликого порішення цих статистичних даних. Проте, результати статистичного моделювання все ще є дуже хорошими, і однозначно підтверджують покращення ефективності побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних. Варто відмітити, що при 10 незалежних змінних, 6 з яких є нулями, 20 експериментальних випробувань

достатньо, щоб застосовувати метод при випадковій величині E з рівномірним розподілом від -4 до 4 . При рівномірному розподілі від -7 до 7 для ефективного застосування досліджуваного методу рекомендується застосовувати 50 експериментальних випробувань. В умовах нормального розподілу випадкової величини E , 20 експериментальних випробувань вистачить для застосування методу при стандартному відхиленні 4 . Сума нормального і рівномірного розподілів випадкової величини E дає результати у зменшенні ефективності роботи, а саме використання 20 експериментальних випробувань вистачить для застосування методу при сумі рівномірного розподілу від -3 до 3 і стандартному відхиленню нормального розподілу 3 . Для ефективного застосування досліджуваного методу при сумі рівномірного розподілу від -6 до 6 та нормального розподілу з стандартним відхиленням 6 випадкової величини E рекомендується застосовувати 50 експериментальних випробувань при 10 вхідних змінних.

5 МАРКЕТИНГОВИЙ АНАЛІЗ СТАРТАП-ПРОЄКТУ

5.1 Опис ідеї проєкту

За результатами проведеного дослідження предметної області було виявлено потребу в створенні програмного забезпечення для статистичного моделювання ефективності оригінального метода побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних, яке б надавало змогу надалі розвивати цей метод за допомогою виявлення його переваг та недоліків, а також надавало б змогу експертам у галузі регресійного аналізу провести свої дослідження і вирішити чи підходить цей метод для розв'язання конкретної задачі. Програмне забезпечення не має аналогів та конкурентів на ринку. У таблицях 5.1 – 5.22 наведено дані, що отримано в процесі маркетингового аналізу стартап-проєкту.

Таблиця 5.1 – Опис ідеї стартап-проєкту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
<p>На основі оригінального методу побудови багатовимірної лінійної регресії, на малому об'ємі експериментальних даних було створене програмне забезпечення, що реалізує цей метод і надає змогу статистично моделювати його ефективність</p>	<p>Подальший розвиток досліджуваного метода за рахунок виявлення його переваг і недоліків.</p>	<p>За допомогою розробленого забезпечення з'являється можливість використовувати і досліджувати оригінальний метод побудови багатовимірної лінійної регресії на малому об'ємі експериментальних даних.</p>
	<p>Аналіз методу зацікавленими сторонами за рахунок проведення власних досліджень, заточених під конкретну доменну область та конкретні значення.</p>	

Таблиця 5.2 – Опис ідеї стартап-проєкту

№	Техніко-економічні характеристики ідеї	Продукція конкурентів			W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проєкт	STATISTICA	SPSS Statistics			
1	Зручність використання	Висока	Висока	Середня			+
2	Підтримка досліджуваного методу	Є	Немає	Немає			+
3	Кросплатформенність	Є	Є	Є		+	
4	Інтеграція з іншими інструментами	Висока	Середня	Середня			+
5	Розширюваність	Висока	Середня	Середня			+

5.2 Технологічний аудит ідеї проєкту

Таблиця 5.3 – Технологічна здійсненність проєкту

№	Ідея проєкту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Бібліотека, що реалізує досліджуваний метод	Python	Наявна	Доступна
2	Система статистичного імітаційного моделювання ефективності досліджуваного методу	Python	Наявна	Доступна
<i>Обрана технологія реалізації ідеї проєкту: Python</i>				

Висновок: технологічна реалізація продукту – можлива, вибрана технологія – мова програмування Python.

5.3 Аналіз ринкових можливостей запуску стартап-проєкту

Таблиця 5.4 – Попередня характеристика потенційного ринку

№	Показники стану ринку	Характеристика
1	Кількість головних гравців, од	2
2	Загальний обсяг продаж, грн./ум.од	900000
3	Динаміка ринку	Зростаюча
4	Наявність обмежень для входу	Високі стандарти безпеки даних при обробці
5	Специфічні вимоги до стандартизації та сертифікації	Немає вимог щодо стандартизації та сертифікації
6	Середня норма рентабельності в галузі або по ринку, %	15-20%

Висновок: виходячи з кількості головних гравців по ринку, зростаючу динаміку ринку, невелику кількість конкурентів та середню норму рентабельності можна зробити висновок, що на даний момент, ринок для входу стартап-продукту є привабливим.

Таблиця 5.5 – Характеристика потенційних клієнтів стартап-проєкту

№	Потреба, що формує ринок	Цільова аудиторія	Відмінності у поведінці цільових груп клієнтів	Вимоги споживачів до товару
1	Можливість використовувати досліджуваний метод	Інженери діагностичних і експертних систем	Можливість вводити свої дані	Споживачі очікують зручний графічний інтерфейс для проведення свого моделювання. Також очікується можливість легко генерувати дані.
2	Можливість надалі розвивати та досліджувати ефективність оригінального методу	Автори методу та особи, які хочуть зробити свій вклад у його розвиток	Можливість генерувати дані	
3	Можливість вивчення роботи досліджуваного методу у конкретній доменній області	Експерти у галузі регресійного аналізу, які досліджують ефективності різних методів побудови БЛР.	Можливість вводити свої дані	

Таблиця 5.6 – Фактори загроз

№	Фактор	Зміст загрози	Можлива реакція компанії
1	Цільова аудиторія	Мала кількість зацікавлених осіб (високий поріг входу)	Розробка аналогічних систем або розширення поточної для дослідження ефективності інших методів для побудови БЛР у професійних та навчальних цілях аби знизити поріг входу
2	Кошти на розробку та підтримку продукту	Закінчення грошей та недостатнє фінансування	Залучення додаткових інвесторів, мотивація роботи на перспективу; Ітеративна розробка продукту задля покрокового виведення продукту на ринок та отримання відповіді користувачів
3	Вихід аналогу	Вихід аналогу даного товару може призвести до знецінення та безідейності даного товару	Вихід товару на ринок в коротші строки з не повною, але достатньою, функціональністю для зацікавлення усіх цільових аудиторій; Проведення рекламної компанії

Таблиця 5.7 – Фактори можливостей

№	Фактор	Зміст можливості	Можлива реакція компанії
1	Новий продукт	Вихід на ринок, Зменшення монополії, Надання нових рішень у сфері	Розробка нової функціональності; Вихід нової продукції на ринок; Надання різноманітних типів ліцензій в залежності від потреб користувача чи замовника.
3	Зворотній зв'язок від користувачів	Можливість отримання необхідної інформації для вдосконалення продукту	Наявність вхідних даних та реакція на них з боку команди розробників задля задоволення потреб та бажань кінцевих користувачів системи кешування даних.
4	Грошова винагорода за рекламу	При достатньому попиту на систему кешування даних можлива комерціалізація продукту на основі реклами задля отримання грошової винагороди для подальшого розвитку продукту	Точкова комерціалізація продукту; Введення реклами; Ведення додаткових коштів у проект задля його подальшого розвитку.

Таблиця 5.8 – Ступеневий аналіз конкуренції на ринку

№	Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1	Тип конкуренції: монополістична	На ринку не існує конкурентів, які дають змогу дослідити ефективність оригінального методу побудови БЛР на малому об'ємі експериментальних даних	Розробка продукту з характеристиками, які покривають сфери вживання що не покривають інші товари-замінники;
2	Рівень конкурентної боротьби: світовий	Всі продукти замінники розроблялись інтернаціональними командами з різних куточків світу, продукти не належать до певної держави, а належать команді розробників	Вихід на ринок збуту продукту з клієнто-необхідною функціональністю; Налагодження маркетингу на основних Інтернет ресурсах задля охоплення великої кількості потенційних користувачів; Надання бета-версій продукту.
3	Галузева ознака: міжгалузева	Конкуренція в інформаційній промисловості, а також з іншими індустріями, як застосовують та досліджують регресійний аналіз	Надання зручного, інтуїтивно зрозумілого інтерфейсу; Підтримка всім відомим методів взаємодії з середовищем розробки; Наявність документації та онлайн підтримки.
4	Конкуренція за видами товарів: товарно-видова	Дана конкуренція – конкуренція між товарами різного виду.	Впровадження функціональності яка відсутня у товарів-замінників.
5	Характер конкурентних переваг: не цінова	Не цінова – надання функціональності, що відсутня у товарах-замінниках.	Впровадження унікальної функціональності.
6	За інтенсивністю: немарочна	Роль торгової марки незначна	Заохочення клієнтів якістю та новизною товару, а не брендом.

Таблиця 5.9 – Аналіз конкуренції в галузі за М. Портером

С	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
к л а д о в і а н а л і з у	Компанії, що надають послуги дослідження методів побудови БЛР	Бар'єри входження високі, треба мати співробітників з високим професійним рівнем та великі інвестиції	Новітні актуальні технології	Бажають знайти універсальний метод для ефективного вирішення задачі побудови багатовимірної лінійної регресії на малому об'ємі даних	Товари-замінники лише дозволять збільшити прогрес через коопераційну стратегію розвитку
В и с н о в к и	Висока здорова конкуренція	Можливості входу високі через велику актуальність технологій	Постачальники не диктують умови, все відбувається завдяки гонці технологій	Існує багато різних методів для побудови БЛР, тому клієнти будуть досліджувати всі наявні і обирати той, який виявиться найбільш дієвим	Товари-замінники дозволять налагодити кооперацію та більш швидкий розвиток нових технологій

Проаналізувавши можливості роботи на ринку з огляду на конкурентну ситуацію можна зробити висновок: оскільки кожний з існуючих продуктів не впливає у великій мірі на поточну ситуацію на ринку в цілому, кожний з існуючих продуктів має свою специфічну сферу використання та свої позитивні та негативні сторони щодо рішення певних типів задач, то робота та вихід на даний ринок є можливою і реалізованою задачею.

Для виходу на ринок продукт повинен мати функціонал що відсутній у продуктів-аналогів, повинен задовольняти потреби користувачів, мати необхідний та достатній функціонал з конфігурування, підтримку зі сторони

розробників та можливість розробки спеціального функціоналу за відповідною ліцензією.

Таблиця 5.10 – Обґрунтування факторів конкурентоспроможності

№	Фактор конкурентоспроможності	Обґрунтування
1	Актуальність товару	Розроблена оригінальна методологія для статистичного імітаційного моделювання ефективності досліджуваного методу
2	Відкритий вихідний код	Можливість розвивати продукт будь-якою людиною у будь-якій точці світу
3	Висока готовність до кооперації	Відкритість до діалогу
4	Простота використання	Інтуїтивно зрозумілий інтерфейс, для використання потрібнет встановлений Python

Таблиця 5.11 – Порівняльний аналіз сильних та слабких сторін системи статистичного імітаційного моделювання ефективності досліджуваного метода

№	Фактор конкурентоспроможності	Б а л л и 1 - 2 0	Рейтинг товарів-конкурентів у порівнянні з запропонованим							
			-3	-2	-1	0	+1	+2	+3	
1	Актуальність товару	20		+						
2	Відкритий вихідний код	19	+							
3	Висока готовність до кооперації	20		+						
4	Простота використання	19		+						

Таблиця 5.12 – SWOT аналіз стартап-проєкту

<p>Сильні сторони (S): Система статистичного моделювання ефективності досліджуваного методу дозволяє визначити оптимальну зону його застосування. На сьогоднішній день не існує подібного програмного продукту.</p>	<p>Слабкі сторони (W): Вихід невідомої компанії на ринок регресійного аналізу. Менший набір функціональності, ніж у конкурентів.</p>
<p>Можливості (O): Розширюваність. Відкритість до діалогу, інтеграції з іншими системами</p>	<p>Загрози (T): Недовіра споживачів. Світова криза. Високий поріг входу. Вихід продукта аналога.</p>

Таблиця 5.13 – Альтернативи ринкового впровадження стартап-проєкту

№	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Безкоштовне надання певного функціоналу у користування споживачам на обмежений термін	Головний ресурс – люди, даний ресурс - наявний	2-3 місяці
2	Реклама	Залучення власних коштів для реклами товару	1-2 місяці
3	Написання статей та опис товару на відомих ресурсах	Головний ресурс – час, даний ресурс - наявний	2-3 тижні
4	Презентація товару на хакатонах й інших ІТ заходах	Ресурс – час та гроші для участі, наявні	1-3 місяці

5.4 Розроблення ринкової стратегії проєкту

Таблиця 5.14 – Вибір цільових груп потенційних споживачів

№	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Вітчизняні інформаційні компанії	Низька, отже вітчизняна галузь ще недостатньо розвинена	10% - низький попит	Низька	Важко через недостатню розвиненість ринку
2	Університети та дослідницькі центри	Висока, то що дуже цінуються новаторство та наукові новинки	90% - високий попит	Висока	Легко, бо дослідницькі установи широко використовують новаторські продукти

Продовження таблиці 5.14

№	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
3	Закордонні інформаційні компанії	Висока, тому що активно використовують схожі інформаційні системи для тестування систем штучного інтелекту	80% - середній попит	Середня	Важко, потрібно розробити англomовну рекламу та документацію
4	Вітчизняні аутсорсингові компанії, що займаються розрyкою інформаційних систем	Висока, тому що активно надають послуги закордонним інформаційним компаніям	90% - середній попит	Середня	Можна увійти у сегмент, зробивши гарну рекламну кампанію

Відповідно до проведеного аналізу можна зробити висновок, що підходящою цільовою групою для розповсюдження даного програмного продукту є працівники ІТ сфери, ІТ компанії в цілому та будь-які підприємства котрі використовують програмні продукти побудовані на мові програмування Java, та використовують реляційні бази даних. Відповідно до стратегії охоплення ринку збуту товару обрано стратегію масового маркетингу, оскільки для підприємств, ІТ працівників та ІТ компаній у цілому надається стандартизований продукт з можливістю розширення функціональності за домовленістю (відповідно до ліцензії).

Таблиця 5.15 – Визначення базової стратегії розвитку

Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
Надання функціональності що відсутня у товарів-замінників, підтримка клієнтів	Проведення реклами, освітлення унікальної функціональності через інтернет ресурси та інші канали, контакт напряду з споживачами; формування лояльності і прихильності споживачів	Прихильність клієнтів; Відмітні властивості товару; Відмітні характеристики товару;	Стратегія концентрованого зростання

Таблиця 5.16 – Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, які?	Стратегія конкурентної поведінки
Так, товар є першопрохідцем, адже наразі немає такого програмного забезпечення, яке дозволяє досліджувати оригінальний метод побудови БЛР	Так, ціль компанії знайти нових споживачів	Ні	Монополія

Таблиця 5.17 – Визначення стратегії позиціонування

№	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проєкту	Вибір асоціацій, які мають сформувати комплексну позицію власного проєкту
1	Можливість заміни різних платформ, що використовує система	Стратегії диверсифікованого зростання	Система буде розроблена з позиції принципів проєктування, що дозволять клієнтам застосовувати різні платформи для інтеграції з уже існуючими продуктами	Кросплатформеність економічна вигода, залучення більшої кількості клієнтів
2	Багатозадачність	Стратегії концентрованого зростання	З самого початку система буде спрямована на використання широкого спектру функцій, який може бути легко доповнений новими	Виконання різних функцій, ефективність, актуальність
3	Здатність системи до розширення	Стратегії диверсифікованого зростання	Система буде побудована модульним способом, що легко дозволить додавати нові модулі без зміни існуючих	Модульність, відкритість до змін, гнучкість

Відповідно до проведеного аналізу можна зробити висновок, що стартап-компанія вибирає як базову стратегію розвитку – стратегію диференціації, як базову стратегію конкурентної поведінки – стратегію заняття конкурентної ніші.

5.5 Розробка маркетингової програми стартап-проекту

Таблиця 5.18 – Визначення ключових переваг концепції потенційного товару

№	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Дослідження ефективності оригінального методу побудови БЛР	Моделювання на великій кількості експериментів та збір статистики	Автоматизоване рішення для багаторазової реалізації оригінального методу для однакових параметрів окрім змінних значень випадкових величин E і збір статистики
2	Використання новітніх рішень розробки ПЗ	Значне пришвидшення роботи програми	Використання бібліотек мови програмування Python

Таблиця 5.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
1. Товар за задумом	Система статистичного імітаційного моделювання ефективності оригінального методу побудови БЛР		
2. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх/Тл/Е/Ор
	Зручний інтерфейс	Нм	Е
	Кросплатформеність	М	Тх
	Легкість у розгортці	М	Тх
	Якість: експериментально доведена відмовостійкість та ефективність		
	Пакування: Характеристики продукту будуть оформлені в офіційній документації на сайті системи, а також на відкритому репозиторії		
	Марка: PaulCorp		

Продовження таблиці 5.19

Рівні товару	Сутність та складові
3. Товар із підкріпленням	До продажу: наявна повна документація, акції на придбання декількох ліцензій, знижки для певних сегментів на покупку товару
	Після продажу: додаткова підтримка спеціалістів налаштування, підтримка з боку розробника
За рахунок чого потенційний товар буде захищено від копіювання: захист інтелектуальної власності, патент	

Таблиця 5.20 – Визначення меж встановлення ціни

Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
Безкоштовні	0 – 100\$	Середній для товару з відкритим кодом і високий для можливої комерційної версії продукту	25 – 40\$ для комерційної версії

Таблиця 5.21 – Формування системи збуту

Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
Цільові клієнти можуть купити підписку на використання застосунку	Аналіз ринку, дослідження ринків для розширення, аналіз законодавчих норм різних країн	Багатоканальний розподіл	Міжнародні торговельні площадки

Таблиця 5.22 – Концепція маркетингових комунікацій

№	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Клієнти зацікавлені особи, що відвідують масові заходи присвячені технологіям, а також особи які активно досліджують область регресійного аналізу	Міжнародні конференції, виставки, саміти, інтернет форуми	Відкритість коду, загальносоціальна мета, зручний інтерфейс для моделювання та роботи з даними	Донести інформацію про можливості та ціль програмного продукту	Зручний застосунок для статистичного імітаційного моделювання ефективності досліджуваного метода

Як результат було створено ринкову (маркетингову) програму, що включає в себе визначення ключових переваг концепції потенційного товару, опис моделі товару, визначення меж встановлення ціни, формування системи збуту та концепцію маркетингових комунікацій.

Висновки до розділу

В п'ятому розділі описано стратегії та підходи з розроблення стартап-проєкту, визначено наявність попиту, динаміку та рентабельність роботи ринку, як висновок було вказано що існує можливість ринкової комерціалізації проєкту. Розглянувши потенційні групи клієнтів, бар'єри входження, стан конкуренції та конкурентоспроможність проєкту було встановлено що проєкт є перспективним. Розглянуто та вибрано альтернативу впровадження стартап-проєкту та доведено доцільність подальшої імплементації проєкту.

ВИСНОВКИ

В дисертаційній роботі розв'язане актуальне наукове завдання підвищення ефективності оцінки коефіцієнтів багатовимірної лінійної регресії на малому об'ємі експериментальних даних, за рахунок створення бібліотеки, що вперше реалізує оригінальний метод побудови багатовимірної лінійної регресії, заданої надлишковим описом, та системи статистичного імітаційного моделювання ефективності цього методу.

У ході виконання магістерської дисертації було проведено критичний аналіз методів та програмного забезпечення, що використовуються для побудови багатовимірної лінійної регресії, заданої надлишковим описом.

На основі даних, отриманих в процесі аналізу, сформульовано задачу підвищення ефективності побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних.

Розроблена оригінальна методологія статистичного імітаційного моделювання ефективності оригінального методу побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних.

Розроблено та обгрунтовано архітектуру бібліотеки, що реалізує оригінальний метод побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних, та застосування, що реалізує оригінальну методологію статистичного імітаційного моделювання його ефективності.

Розроблено програмне забезпечення, що включає в себе кросплатформенну бібліотеку, що реалізує оригінальний алгоритм побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних, та систему статистичного імітаційного моделювання ефективності застосування досліджуваного методу.

Для розробки програмних продуктів було використано мову програмування Python, що дозволяє зручно та швидко інтегрувати розроблену бібліотеку в програмне забезпечення користувача.

За допомогою розробленої системи статистичного імітаційного моделювання досліджуваного метода визначено рекомендації щодо його ефективного використання.

Результати роботи над магістерською дисертацією опубліковані у статті.

Наукова новизна одержаних результатів магістерської дисертації:

- розроблена оригінальна методологія проведення статистичних імітаційних експериментів для обґрунтування області ефективного використання універсального методу побудови багатовимірних лінійних регресій, заданих надлишковим описом, по невеликому об'єму експериментальних даних;

- розроблена та обґрунтована структура програмного забезпечення, що вперше реалізує оригінальний метод побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних та систему статистичного імітаційного моделювання для знаходження області його ефективного використання.

Практична значимість одержаних результатів полягає у розробці кросплатформенної бібліотеки (моноліт), що реалізує оригінальний алгоритм побудови багатовимірної лінійної регресії, заданої надлишковим описом, на малому об'ємі експериментальних даних, яка може використовуватись у сучасних діагностичних і експертних системах.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- 1) Abdulrahman A.T., Alshammari N.S. Factor analysis and regression analysis to find out the influencing factors that led to the countries' debt crisis. *Advances and Applications in Statistics* 78, 1–16 (2022).
- 2) Agresti A. *Categorical Data Analysis*. Wiley-Interscience, New York (2002).
- 3) Babatunde, G., et al. Impact of climatic change on agricultural product yield using k-means and multiple linear regressions. *International Journal of Education and Man-agement Engineering* 9 (3), 16–26 (2019).
- 4) Braak C.J.F. ter, Looman C.W.N. Regression. In: Jongman, R.H.G., ter Braak, C.J.F., van Tongeren, O.F.R. *Data analysis in community and landscape ecology*, pp. 29–77. Cambridge University Press, Cambridge, 1995.
- 5) Draper N.R., Smith H. *Applied Regression Analysis*, 3rd edn. Wiley & Sons, New York (1998).
- 6) Flitman A.M. Towards analysing student failures: neural networks compared with regression analysis and multiple discriminant analysis. *Computers & Operations Re-search* 24 (4), 367–377 (1997).
- 7) Hudson D.J. *Statistics Lectures, Volume 2: Maximum Likelihood and Least Squares Theory*. CERN Reports 64(18). CERN, Geneva (1964).
- 8) Івахенко А. Г. *Моделювання складних систем*. Вища школа, Київ(1987).
- 9) Johnson R.A., Wichern D.W. *Applied Multivariate Statistical Analysis*, 5th edn. Prentice-Hall, Upper Saddle River, NJ (2002).
- 10) Knowles D., Parts L., Glass D., Winn J.M. Modeling skin and ageing phenotypes using latent variable models in infer.net. Paper presented at Predictive Models in Personalized Medicine Workshop NIPS 2010, Vancouver, 6-11 Dec 2010.
- 11) Lio W., Liu B. Uncertain maximum likelihood estimation with application to uncertain regression analysis. *Soft Computing* 24 (13), 9351–9360.
- 12) Liu S.S., Zhu, Y. Simultaneous Maximum Likelihood Estimation for Piecewise Linear Instrumental Variable Models. *Entropy* 24 (9), 1235 (2022).

13) Mallet A.A. A maximum likelihood estimation method for random coefficient re-gression models. *Biometrika* 73 (3), 645–656 (1986).

14) Настенко Е., Павлов О., Бойко Г., Носовець О. Многокритеріальний алгоритм шарової регресії. *Біомедична інженерія і технологія* 2020 (3), 48–53.

15) Pavlov A., Holovchenko M., Mukha I., et al. Mathematics and software for build-ing nonlinear polynomial regressions using estimates for univariate polynomial re-gressions coefficients with a given (small) variance. In: *Advances in Computer Sci-ence for Engineering and Education V. ICCSSEA 2022. Lecture Notes on Data En-gineering and Communications Technologies* 134, 288–303 (2022).

16) Павлов О. А., Головченко М. М., Ревіч М. Метод оцінки коефіцієнтів при лінійних членах багатовимірної поліноміальної регресії, заданої надлишковим описом. *Адаптивні системи автоматичного управління міжвідомчій науково тухнічний збірник*. Київ: НТУУ «КПІ». 1 (40), 110–117 (2022).

17) Павлов О. А., Головченко М. М. Побудова одновимірної і багатовимірної поліноміальної регресії за надлишковим описом з використанням активного експерименту. *Вісник Національного технічного університету «ХПІ» Серія: системний аналіз, у правління та інформаційні технології* 1 (3), 3–8 (2020).

18) Павлов О. А., Чеховський А. В. Модифікація алгоритму побудови багатовимірної поліноміальної регресії за надлишковим описом. *Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології*. 29, 50–54 (2012).

19) Павлов О. А., Чеховський А. В. Побудова багатовимірної поліноміальної регресії. *Активний експеримент з обмеженнями*. Інформаційний вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології. 2009 (4), 174–185 (2009).

20) Павлов О. А., Головченко М. М., Дрозд В. В. Побудова багатовимірного полінома, заданого надлишковим описом в стохастичній та детермінованій постановках з використанням активного експерименту.

Інформаційний вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, контроль та інформаційні технології. 1 (7), 3–8 (2022).

21) Ruff L. et al. Deep one-class classification. Paper presented at 35th International Conference on Machine Learning, 80th edn. PMLR, 4393–4402 (2018).

22) Scikit-learn Library: Machine learning in Python [Електронний ресурс] // Режим доступу: <https://scikit-learn.org/stable/>

23) Scott J.T. Factor Analysis and Regression. *Econometrica* 34 (3), 552–562 (1966).

24) SPSS Statistics [Електронний ресурс] // Режим доступу: <https://www.ibm.com/products/spss-statistics>

25) TIBCO Statistica 14.1.0 [Електронний ресурс] // Режим доступу: <https://docs.tibco.com/products/tibco-statistica-14-1-0>

26) Sur P., Candès E. J. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences* 116 (29), 14516–14525 (2019).

27) The R Manuals [Електронний ресурс] // Режим доступу: <https://cran.r-project.org/manuals.html>

28) Згуровський М. З., Павлов О.А. Застосування рішень в мережевих системах з обмеженими ресурсами. Наукова думка, Київ (2010).

29) Згуровський М. З., Павлов О. А. Складні задачі комбінаторної оптимізації у плануванні та прийнятті. Наукова думка, Київ (2016).

30) Павлов О. А., Головченко М. М. Модифікований метод побудови багатовимірної лінійної регресії, заданої надлишковим описом. Київ (2024).

ДОДАТКИ

ДОДАТОК А

Діаграма архітектури

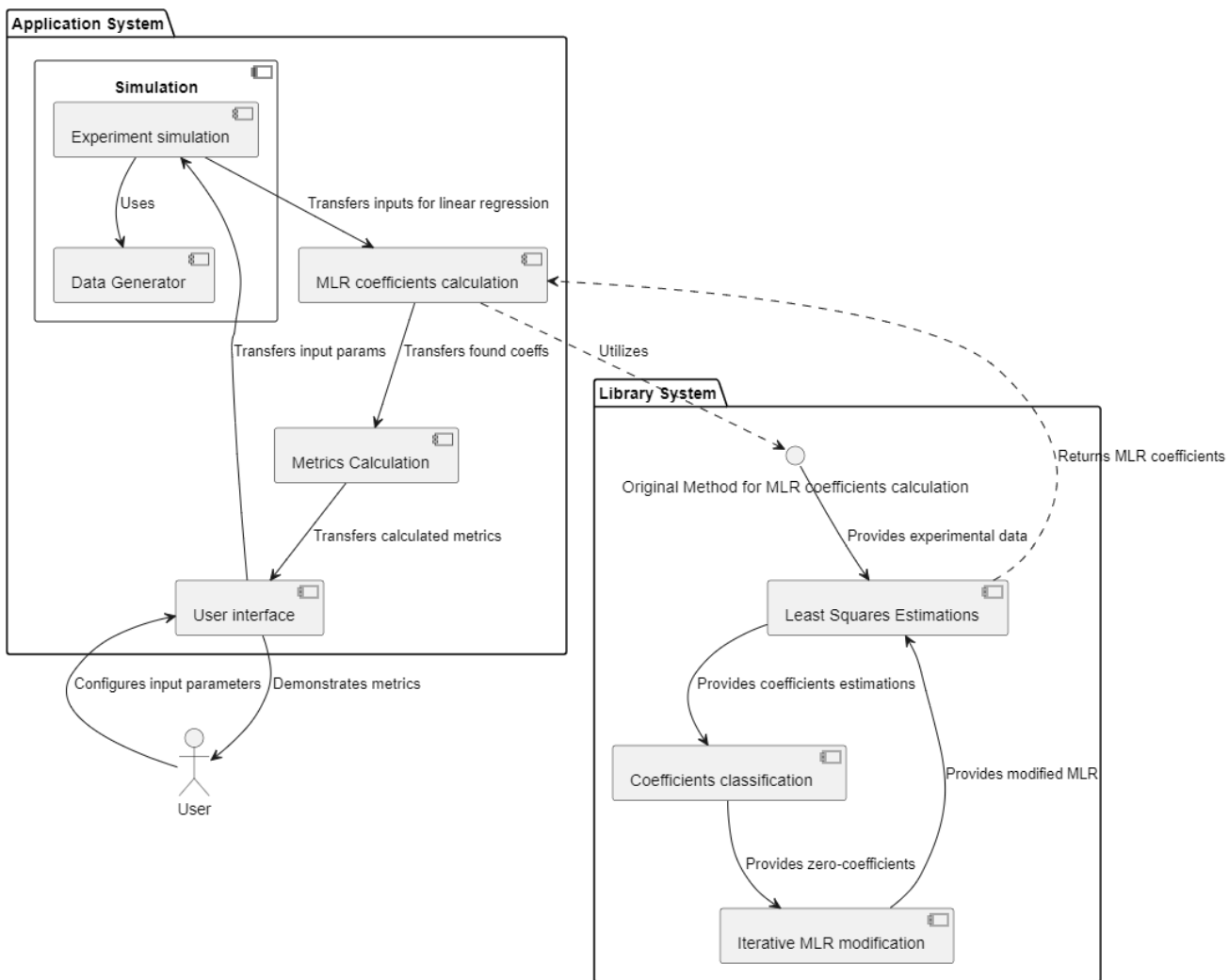


Рисунок А.1 – Діаграма архітектури

ДОДАТОК Б

Діаграма класів

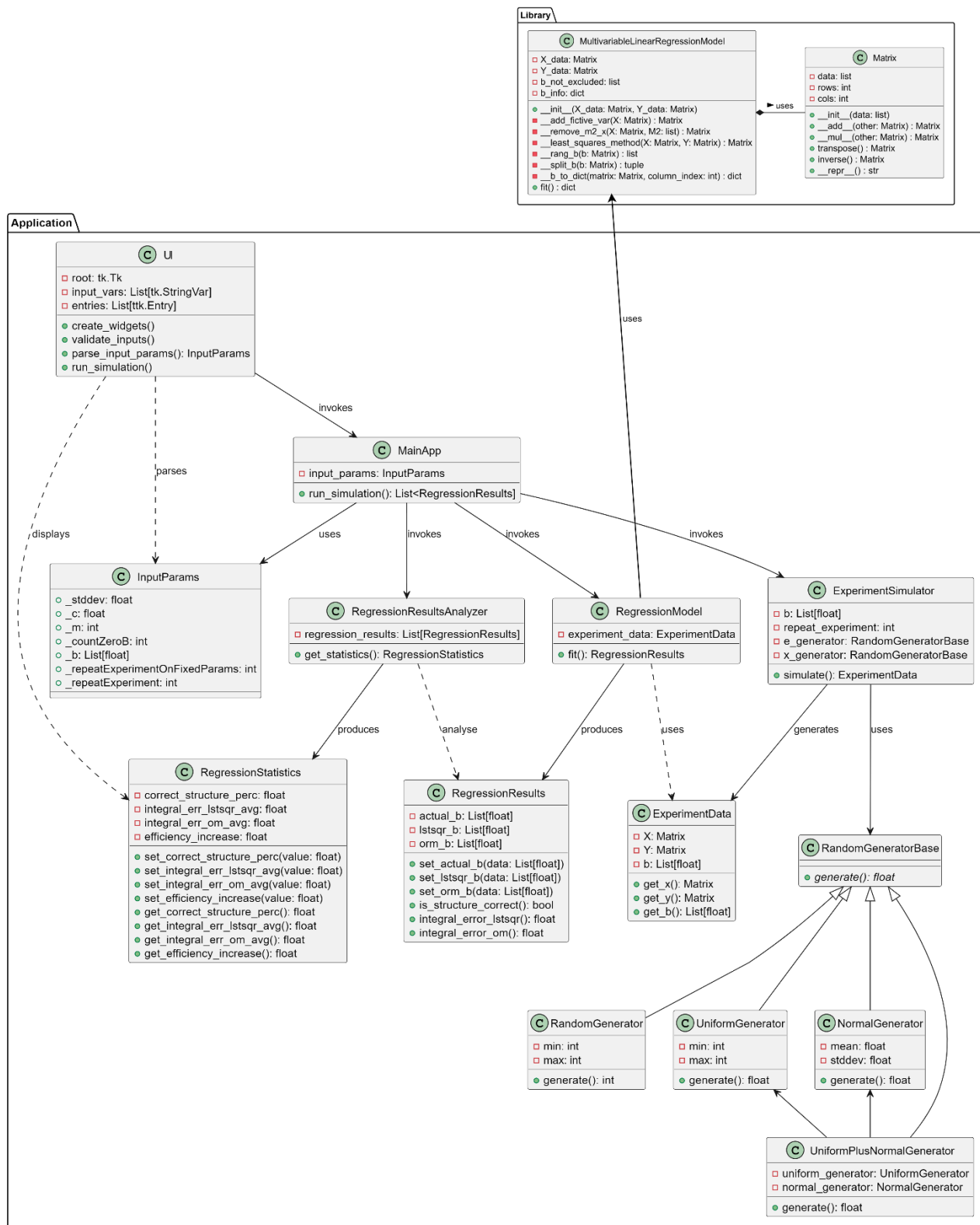


Рисунок Б.1 – Діаграма класів

ДОДАТОК В

Лістинг коду

Файл `effectivness_research_system.py`

```
import tkinter as tk
from tkinter import messagebox
from tkinter import ttk, font # ttk for better styling
from app import InputParams, MainApp
from results_analyzer import RegressionResultsAnalyzer

class InputUI:
    def __init__(self, master):
        self.master = master
        # self.master.resizable(False, False)
        self.master.title("Application")
        self.master.configure(bg="#f5f5f5")
        self.master.bind("<Configure>", self.adjust_wrap_length) # Bind adjust on resize
        self.dist_var = tk.IntVar(value=1)
        self.dist_widget = []
        self.dist_widget_row = None
        self.list_b_entries = []
        self.scrollbar_row = None
        self.red_style = ttk.Style()
        self.red_style.configure("Red.TEntry", fieldbackground='red', foreground='white', background='red')
        self.m_var = tk.StringVar(value="10")
        self.zero_b_var = tk.StringVar(value="5")
        self.exp_count_var = tk.StringVar(value="20")
        self.repeat_exp_var = tk.StringVar(value="10")
        self.c_var = tk.IntVar(value="2.0")
        self.stddev_var = tk.StringVar(value="1.0")
        self.b_vars = []
        self.create_widgets()

    def create_widgets(self):
        title_font = font.Font(family="Times New Roman", size=18, weight="bold")
        label_font = font.Font(family="Times New Roman", size=14)
        grid_row = 0
        # Initial wraplength can be roughly the screen width
        self.title_label = ttk.Label(self.master, text="Simulation of Original Method for Constructing Multivariable Linear Regression on Small Amount of Experimental Data",
            font=title_font, wraplength=self.master.winfo_width() - 160, justify="center")
        self.title_label.grid(row=grid_row, column=0, columnspan=4, sticky="ew", pady=20, padx=(80, 70))
        grid_row += 1
        self.master.grid_columnconfigure(0, weight=1)

        self.single_widgets_frame = tk.Frame(self.master)
        self.single_widgets_frame.grid(row=grid_row, column=0, columnspan=3, pady=0)
        pady = 10
        # m
        ttk.Label(self.single_widgets_frame, text="Number of independent values (m, 1..30):", font=label_font, justify=tk.RIGHT).grid(row=grid_row,
            column=0, pady=pady, padx=40, sticky="w")
        self.m_entry = tk.Entry(self.single_widgets_frame, textvariable=self.m_var)
        self.m_entry.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
        self.m_var.trace_add("write", self.update_list_b_entries)
        grid_row += 1
        self.master.grid_columnconfigure(0, weight=1)

        # zero coefficients
        ttk.Label(self.single_widgets_frame, text="Number of zero coefficients (1..m):", font=label_font, justify=tk.RIGHT).grid(row=grid_row,
            column=0, pady=pady, padx=40, sticky="w")
        self.entry_zero_coefficients_count = tk.Entry(self.single_widgets_frame, textvariable=self.zero_b_var)
        self.entry_zero_coefficients_count.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
        self.zero_b_var.trace_add("write", self.validate_zero_b)
        grid_row += 1
        self.master.grid_columnconfigure(0, weight=1)

        # amount of data
        ttk.Label(self.single_widgets_frame, text="Number of experiments:", font=label_font, justify=tk.RIGHT).grid(row=grid_row, column=0,
            pady=pady, padx=40, sticky="w")
        self.entry_experiments_amount = tk.Entry(self.single_widgets_frame, textvariable=self.exp_count_var)
        self.entry_experiments_amount.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
        self.exp_count_var.trace_add("write", self.validate_exp_count_var)
        grid_row += 1
        self.master.grid_columnconfigure(0, weight=1)
```

```

# the number of repetitions of the experiment
tk.Label(self.single_widgets_frame, text="Number of repetitions of the experiment (10..1000):", font=label_font,
justify=tk.RIGHT).grid(row=grid_row, column=0, pady=pady, padx=40, sticky="w")
self.entry_repeat_experiment_count = tk.Entry(self.single_widgets_frame, textvariable=self.repeat_exp_var)
self.entry_repeat_experiment_count.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
self.repeat_exp_var.trace_add("write", self.validate_repeat_exp_var)
grid_row += 1
self.master.grid_columnconfigure(0, weight=1)

tk.Label(self.single_widgets_frame, text="E-distribution:", font=label_font, justify=tk.RIGHT).grid(row=grid_row, column=0, pady=pady,
padx=40, sticky="w")
self.radio_frame = tk.Frame(self.single_widgets_frame)
tk.Radiobutton(self.radio_frame, text="Uniform", variable=self.dist_var, value=1, command=self.show_e_fields).grid(row=0, column=1,
pady=0, padx=10, sticky="ew")
tk.Radiobutton(self.radio_frame, text="Normal", variable=self.dist_var, value=2, command=self.show_e_fields).grid(row=0, column=2, pady=0,
padx=10, sticky="ew")
tk.Radiobutton(self.radio_frame, text="Uniform + Normal", variable=self.dist_var, value=3, command=self.show_e_fields).grid(row=0,
column=3, pady=0, padx=10, sticky="ew")
self.radio_frame.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
grid_row += 1
self.dist_widget_row = grid_row

self.uniform_distr_label = tk.Label(self.single_widgets_frame, text="Uniform distributions parameter (1..10):", font=label_font,
justify=tk.RIGHT)
self.uniform_distr_label.grid(row=grid_row, column=0, pady=pady, padx=40, sticky="w")
self.uniform_distr_param_entry = tk.Entry(self.single_widgets_frame, textvariable=self.c_var)
self.uniform_distr_param_entry.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
self.c_var.trace_add("write", self.validate_c_var)
grid_row += 1
self.master.grid_columnconfigure(0, weight=1)

self.normal_distr_label = tk.Label(self.single_widgets_frame, text="Normal distributions standart deviation (1..10):", font=label_font,
justify=tk.RIGHT)
# self.normal_distr_label.grid(row=grid_row, column=0, pady=0, padx=40, sticky="w")
self.normal_distr_param_entry = tk.Entry(self.single_widgets_frame, textvariable=self.stddev_var)
self.stddev_var.trace_add("write", self.validate_stddev_var)
# self.normal_distr_param_entry.grid(row=grid_row, column=1, pady=0, padx=40, sticky="ew")
grid_row += 1
self.master.grid_columnconfigure(0, weight=1)

self.b_label = tk.Label(self.master, text="MLR coefficients:", font=label_font, justify="center")

self.b_frame = tk.Frame(self.master)
self.b_frame.grid(row=4, column=0, sticky="ew")
self.b_frame.grid_columnconfigure(0, weight=1)
self.b_frame.grid_rowconfigure(0, weight=1)
self.b_frame.grid_propagate(False)

self.canvas = tk.Canvas(self.b_frame)
self.canvas.grid(row=0, column=0, sticky="news", padx=40)

self.scrollbar = tk.Scrollbar(self.b_frame, orient="horizontal", command=self.canvas.xview)
self.scrollbar.grid(row=1, column=0, sticky="ew")
self.canvas.configure(xscrollcommand=self.scrollbar.set)

self.scrollable_frame = tk.Frame(self.canvas)
self.canvas.create_window((0, 0), window=self.scrollable_frame, anchor="nw")

# Button that triggers submit functionality
button_font = font.Font(family="Helvetica", size=16, weight="bold")

# Button that triggers submit functionality with the specified font
self.submit_button = tk.Button(self.master, text="Run experiments", command=self.pipeline, style="Custom.TButton")
self.submit_button.grid(row=5, column=0, pady=20)

self.update_list_b_entries()

# Apply a style that uses the custom font
style = ttk.Style()
style.configure("Custom.TButton", font=button_font)

def validate_zero_b(self, *args):
self.entry_zero_coefficients_count.config(foreground='black')
try:
# add check
v = int(self.zero_b_var.get())
if(v < 1):
self.entry_zero_coefficients_count.config(foreground='red')
return

```

```

except ValueError:
    self.entry_zero_coefficients_count.config(background='red')

def validate_exp_count_var(self, *args):
    self.entry_experiments_amount.config(background='black')
    try:
        # add check
        v = int(self.exp_count_var.get())
        if(v < 1):
            self.entry_experiments_amount.config(background='red')
            return
    except ValueError:
        self.entry_experiments_amount.config(background='red')

def validate_repeat_exp_var(self, *args):
    self.entry_repeat_experiment_count.config(background='black')
    try:
        v = int(self.repeat_exp_var.get())
        if(v < 10 or v > 1000):
            self.entry_repeat_experiment_count.config(background='red')
            return
    except ValueError:
        self.entry_repeat_experiment_count.config(background='red')

def validate_c_var(self, *args):
    self.uniform_distr_param_entry.config(background='black')
    try:
        v = float(self.c_var.get())
        if(v > 10.0 or v < 1.0):
            self.uniform_distr_param_entry.config(background='red')
            return
    except ValueError:
        self.uniform_distr_param_entry.config(background='red')

def validate_stddev_var(self, *args):
    self.normal_distr_param_entry.config(background='black')
    try:
        v = float(self.stddev_var.get())
        if(v > 10.0 or v < 1.0):
            self.normal_distr_param_entry.config(background='red')
            return
    except ValueError:
        self.normal_distr_param_entry.config(background='red')

def count_zero_b(self):
    count = 0
    for b in self.b_vars:
        if(b.get() == "0"):
            count += 1
    return count

def validate_b_var(self, i, *args):
    self.list_b_entries[i].config(background='black')
    try:
        zero_c = self.get_zero_b()
        if (zero_c != self.count_zero_b()):
            for j in range(len(self.list_b_entries)):
                self.list_b_entries[j].config(background='blue')
        else:
            for j in range(len(self.list_b_entries)):
                self.list_b_entries[j].config(background='black')
        v = float(self.b_vars[i].get())
    except ValueError:
        self.list_b_entries[i].config(background='red')

def update_list_b_entries(self, *args):
    self.m_entry.config(background='black')
    try:
        m = int(self.m_var.get())
        if(m < 1 or m > 40):
            self.m_entry.config(background='red')
            return
        for widget in self.scrollable_frame.winfo_children():
            widget.destroy()
        self.list_b_entries = []
        self.b_vars = []
        self.b_label.grid(row=3, column=0, padx=40, pady=(20,0))
        for i in range(m + 1):
            placeholder_text = f"b{i}"

```

```

self.b_vars.append(tk.StringVar(value="0"))
tk.Label(self.scrollable_frame, text=placeholder_text, width=10).grid(row=1, column=i, sticky='news')
entry = ttk.Entry(self.scrollable_frame, textvariable=self.b_vars[i], width=8)
self.b_vars[i].trace_add("write", lambda *args, id=i: self.validate_b_var(id, *args))
entry.grid(row=2, column=i, sticky='news')

self.list_b_entries.append(entry)
self.scrollable_frame.update_idletasks()
self.b_frame.config(width=800, height=self.scrollbar.winfo_height()+43)

self.canvas.config(scrollregion=self.canvas.bbox("all"))

except ValueError:
    self.m_entry.config(foreground='red')

def get_stddev(self):
    if(self.dist_var.get() == 1):
        return 0.0
    return float(self.stddev_var.get())

def get_c(self):
    if(self.dist_var.get() == 2):
        return 0.0
    return float(self.c_var.get())

def get_m(self):
    return int(self.m_var.get())

def get_zero_b(self):
    return int(self.zero_b_var.get())

def get_b(self):
    b = []
    for b_s in self.b_vars:
        b.append(float(b_s.get()))
    return b

def exp_count(self):
    return int(self.exp_count_var.get())

def repeat_exp(self):
    return int(self.repeat_exp_var.get())

def parse_inputs(self) -> InputParams:
    return InputParams(self.get_stddev(), self.get_c(), self.get_m(), self.get_zero_b(), self.get_b(), self.exp_count(), self.repeat_exp())

def pipeline(self):
    # Placeholder function to demonstrate responses to clicking the Submit button
    try:
        p = self.parse_inputs()
        app = MainApp(p)
        results = app.run_simulation()
        stats = RegressionResultsAnalyzer(results).get_statistics()

        self.open_new_window(stats)
    except ValueError as e:
        messagebox.showerror("Error", f'An error occurred: {e}')

def open_new_window(self, stats):
    title_font = font.Font(family="Times New Roman", size=18, weight="bold")
    label_font = font.Font(family="Times New Roman", size=14)
    new_window = tk.Toplevel(self.master)
    new_window.title("New Window")
    new_window.geometry("600x300")

    label = tk.Label(new_window, text="Experiment results:", font=title_font)
    label.grid(row=0, column=0, padx=20, pady=10)
    perc = stats.get_correct_structure_perc()*100
    avg_lst = 0
    if stats.get_integral_err_lstsq_avg():
        avg_lst = stats.get_integral_err_lstsq_avg()
    avg_om = 0
    if stats.get_integral_err_om_avg():
        avg_om = stats.get_integral_err_om_avg()

    eff_inc = 0
    if stats.get_efficiency_increase():
        eff_inc = stats.get_efficiency_increase()*100

```

```

output_frame = ttk.Frame(new_window)
output_frame.grid(row=1, column=0, columnspan=2, padx=20)
tk.Label(output_frame, text="Percentage of correctly found MLR structures", font=label_font).grid(row=0, column=0, padx=20, pady=10)
tk.Label(output_frame, text=f"{perc:.2f}%", font=label_font).grid(row=0, column=1, padx=20, pady=10)
tk.Label(output_frame, text="Avarage integral error while using least squares", font=label_font).grid(row=1, column=0, padx=20, pady=10)
tk.Label(output_frame, text=f"{avg_lst:.3f}", font=label_font).grid(row=1, column=1, padx=20, pady=10)
tk.Label(output_frame, text="Avarage integral error while using original method", font=label_font).grid(row=2, column=0, padx=20, pady=10)
tk.Label(output_frame, text=f"{avg_om:.3f}", font=label_font).grid(row=2, column=1, padx=20, pady=10)
tk.Label(output_frame, text="Percent improvement in least squares efficiency", font=label_font).grid(row=3, column=0, padx=20, pady=10)
tk.Label(output_frame, text=f"{eff_inc:.2f}%", font=label_font).grid(row=3, column=1, padx=20, pady=10)

def show_e_fields(self):
    grid_row = self.dist_widget_row
    pady = 10
    self.uniform_distr_label.grid_remove()
    self.uniform_distr_param_entry.grid_remove()
    self.normal_distr_label.grid_remove()
    self.normal_distr_param_entry.grid_remove()
    if self.dist_var.get() == 1:
        self.uniform_distr_label.grid(row=grid_row, column=0, pady=pady, padx=40, sticky="w")
        self.uniform_distr_param_entry.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
        self.master.grid_columnconfigure(0, weight=1)
    elif self.dist_var.get() == 2:
        self.normal_distr_label.grid(row=grid_row, column=0, pady=pady, padx=40, sticky="w")
        self.normal_distr_param_entry.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
        self.master.grid_columnconfigure(0, weight=1)
    elif self.dist_var.get() == 3:
        self.uniform_distr_label.grid(row=grid_row, column=0, pady=pady, padx=40, sticky="w")
        self.uniform_distr_param_entry.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
        self.master.grid_columnconfigure(0, weight=1)
        grid_row += 1
        self.normal_distr_label.grid(row=grid_row, column=0, pady=pady, padx=40, sticky="w")
        self.normal_distr_param_entry.grid(row=grid_row, column=1, pady=pady, padx=40, sticky="ew")
        self.master.grid_columnconfigure(0, weight=1)

def adjust_wrap_length(self, event=None):
    # Update wraplength to the width of the window minus a small padding
    current_width = self.master.winfo_width()
    self.title_label.configure(wraplength=current_width - 80)

class Application:
    def __init__(self, root):
        self.ui = InputUI(root)

def main():
    root = tk.Tk()
    root.geometry("900x600")
    app = Application(root)
    root.mainloop()

if __name__ == "__main__":
    main()

```

Файл experiment_simulator.py

```

from typing import List
import numpy as np
import random
import math
import csv
from original_least_squares_boost.original_least_squares_boost import MultivariableLinearRegressionModel
from random_generators import RandomGeneratorBase, RandomGenerator, UniformGenerator, NormalGenerator, UniformPlusNormalGenerator
from matrix import Matrix

class ExperimentData:
    def __init__(self, X, Y, b):
        self.X = X
        self.Y = Y
        self.b = b

    def get_x(self):
        return self.X

    def get_y(self):
        return self.Y

    def get_b(self):
        return self.b

```

```

class ExperimentSimulator:
    def __init__(self, b, repeat_experiment: int, E_generator: RandomGeneratorBase, X_generator = UniformGenerator(1, 20)):
        self.b = b
        self.m = len(b)
        self.repeat_experiment = repeat_experiment
        self.e_generator = E_generator
        self.x_generator = X_generator

    def simulate(self) -> ExperimentData:
        X = self.__generate_x()
        # print("x")
        # print(X)
        Y = self.__calculate_y(X)
        # print("y")
        # print(Y)
        return ExperimentData(X, Y, self.b)

    def __generate_x(self):
        data = np.zeros((self.repeat_experiment, self.m))
        data[:, 0] = 1
        for i in range(1, self.m):
            for j in range(0, self.repeat_experiment):
                data[j, i] = self.x_generator.generate()
        return Matrix(data)

    def __shum_y(self, Y: Matrix):
        newY = Y.data
        for i in range(0, len(Y.data)):
            newY[i][0] = Y.data[i][0] + self.e_generator.generate()
        return Matrix(newY)

    def __calculate_y(self, X: Matrix):
        B = Matrix(np.array(self.b).reshape(-1, 1)) # Transforming list to column matrix
        Y = X * B
        return self.__shum_y(Y)

```

Файл main_app.py

```

from typing import List
import numpy as np
import random
import math
import csv
from original_least_squares_boost.original_least_squares_boost import MultivariableLinearRegressionModel
from random_generators import UniformPlusNormalGenerator
from matrix import Matrix
from experiment_simulator import ExperimentSimulator, ExperimentData
from regression_model import RegressionModel, RegressionResults
from results_analyzer import RegressionResultsAnalyzer

```

```

class InputParams:
    def __init__(self, stddev: float, c: float, m: int, countZeroB: int, b: List[float], repeatExperimentOnFixedParams: int, repeatExperiment: int):
        self._stddev = stddev
        self._c = c
        self._m = m
        self._countZeroB = countZeroB
        self._b = b
        self._repeatExperimentOnFixedParams = repeatExperimentOnFixedParams
        self._repeatExperiment = repeatExperiment

```

```

class MainApp:
    def __init__(self, input_params: InputParams):
        self._input_params = input_params
        print("Input parameters:")
        print("Dimensionality of BLR:\t\t\t", self._input_params._m)
        print("Number of zero coefficients:\t\t\t", self._input_params._countZeroB)
        print("Values of non-zero BLR coefficients:\t\t\t", ''.join(map(str, self._input_params._b)))
        print("Repetitions with fixed parameters:\t\t\t", self._input_params._repeatExperimentOnFixedParams)
        print("Repetitions with the given simulation parameters:\t\t", self._input_params._repeatExperiment)
        print("The variance of the normal distribution E:\t\t", self._input_params._stddev)
        print("Uniform distribution parameter E:\t\t\t", self._input_params._c)

    def run_simulation(self):
        results = []
        for a in range(self._input_params._repeatExperiment):

```

```

        simulator = ExperimentSimulator(self._input_params._b, self._input_params._repeatExperimentOnFixedParams,
        UniformPlusNormalGenerator(-self._input_params._c, self._input_params._c, 0.0, self._input_params._stddev))
        experiment_data = simulator.simulate()
        regression_model = RegressionModel(experiment_data)
        regression_results = regression_model.fit()
        # print(regression_results)
        results.append(regression_results)
    return results

```

Файл matrix.py

```

class Matrix:
    def __init__(self, data):
        self.data = data
        self.rows = len(data)
        self.cols = len(data[0])

    def __add__(self, other):
        if self.rows != other.rows or self.cols != other.cols:
            raise ValueError("Matrices are not the same size.")

        result = [[self.data[i][j] + other.data[i][j] for j in range(self.cols)]
                  for i in range(self.rows)]
        return Matrix(result)

    def __mul__(self, other):
        if self.cols != other.rows:
            raise ValueError("Matrix multiplication not possible: Column of A does not match row of B")

        result = [[sum(a * b for a, b in zip(row_a, col_b))
                  for col_b in zip(*other.data)] for row_a in self.data]
        return Matrix(result)

    def transpose(self):
        result = [[self.data[i][j] for i in range(self.rows)]
                  for j in range(self.cols)]
        return Matrix(result)

    def inverse(self):
        from numpy.linalg import inv, LinAlgError
        try:
            inv_matrix = inv(self.data)
        except LinAlgError:
            raise ValueError("Matrix is not invertible")
        return Matrix(inv_matrix)

    def __repr__(self):
        return "\n".join(["\t".join(map(str, row)) for row in self.data])

```

Файл random_generators.py

```

import numpy as np
from abc import ABC, abstractmethod

class RandomGeneratorBase(ABC):
    """Abstract base class for random number generation."""

    @abstractmethod
    def generate(self):
        """Generate a random number."""
        pass

class RandomGenerator(RandomGeneratorBase):
    """Generate a pseudorandom number between lower and upper bounds (inclusive). """
    def __init__(self, min, max):
        self.min = min
        self.max = max
        if min > max:
            raise ValueError("Lower bound must be less than or equal to upper bound.")

    def generate(self):
        return np.random.randint(self.min, self.max)

class UniformGenerator(RandomGeneratorBase):
    """Generate a uniformly distributed random number between lower and upper bounds. """
    def __init__(self, min, max):
        self.min = min
        self.max = max

```

```

    if min > max:
        raise ValueError("Lower bound must be less than or equal to upper bound.")

    def generate(self):
        return np.random.uniform(self.min, self.max)

class NormalGenerator(RandomGeneratorBase):
    def __init__(self, mean, stddev):
        self.mean = mean
        self.stddev = stddev

    def generate(self):
        return np.random.normal(self.mean, self.stddev)

class UniformPlusNormalGenerator(RandomGeneratorBase):
    def __init__(self, uniform_min, uniform_max, normal_mean, normal_stddev):
        self.uniform_generator = UniformGenerator(uniform_min, uniform_max)
        self.normal_generator = NormalGenerator(normal_mean, normal_stddev)

    def generate(self):
        return self.uniform_generator.generate() + self.normal_generator.generate()

```

Файл regression_model.py

```

from typing import List
import numpy as np
import random
import math
import csv
from original_least_squares_boost.original_least_squares_boost import MultivariableLinearRegressionModel
from random_generators import UniformPlusNormalGenerator
from matrix import Matrix
from experiment_simulator import ExperimentSimulator, ExperimentData

def avg_sqrt_square_of_elements(elements):
    avg = math.sqrt(sum([(el ** 2) for el in elements]))
    return avg

def normalize_by_avg_sqrt_square(elements):
    avg = avg_sqrt_square_of_elements(elements)
    normalized_list = [el / avg for el in elements]
    return normalized_list

def subtract_lists(l1, l2):
    if len(l1) != len(l2):
        raise ValueError("Lists must have the same length.")
    result = [l1[i] - l2[i] for i in range(len(l1))]
    return result

def integral_error(l1, l2):
    norm_l1 = normalize_by_avg_sqrt_square(l1)
    norm_l2 = normalize_by_avg_sqrt_square(l2)
    return avg_sqrt_square_of_elements(subtract_lists(norm_l1, norm_l2))

class RegressionResults:
    def __init__(self, actual_b=None, lstsq_b=None, orm_b=None):
        self.actual_b = actual_b if actual_b is not None else []
        self.lstsq_b = lstsq_b if lstsq_b is not None else []
        self.orm_b = orm_b if orm_b is not None else []

    def set_actual_b(self, data):
        self.actual_b = data

    def set_lstsq_b(self, data):
        self.lstsq_b = data

    def set_orm_b(self, data):
        self.orm_b = data

    def is_structure_correct(self):
        if len(self.actual_b) != len(self.orm_b):
            return False
        for i in range(len(self.actual_b)):
            if((self.actual_b[i] == 0 or self.orm_b[i] == 0) and self.actual_b[i] != self.orm_b[i]):
                return False
        return True

    def integral_error_lstsq(self):

```

```

    return integral_error(self.actual_b, self.lstsq_b)

def integral_error_om(self):
    return integral_error(self.actual_b, self.orm_b)

def __str__(self):
    return (f"\nactual_b: {self.actual_b}\n"
            f"lstsq_b: {self.lstsq_b}\n"
            f"orm_b: {self.orm_b}\n"
            f"correct: {self.is_structure_correct()}\n"
            f"lstsq err: {self.integral_error_lstsq()}\n"
            f"omg err: {self.integral_error_om()}\n")

def least_squares_method(X: Matrix, Y: Matrix):
    b, residuals, rank, s = np.linalg.lstsq(X.data, Y.data, rcond=None)
    return b

class RegressionModel:
    def __init__(self, experiment_data: ExperimentData):
        self._experiment_data = experiment_data

    def fit(self) -> RegressionResults:
        lstsq_b = least_squares_method(self._experiment_data.get_x(), self._experiment_data.get_y())
        om = MultivariableLinearRegressionModel(self._experiment_data.get_x().data, self._experiment_data.get_y().data)
        om_b = om.fit()
        result = RegressionResults(self._experiment_data.get_b(), [item[0] for item in lstsq_b], list(om_b.values()))
        return result

```

Файл results_analyzer.py

```

from typing import List
import numpy as np
import random
import math
import csv
from original_least_squares_boost.original_least_squares_boost import MultivariableLinearRegressionModel
from random_generators import UniformPlusNormalGenerator
from matrix import Matrix
from experiment_simulator import ExperimentSimulator, ExperimentData
from regression_model import RegressionModel, RegressionResults

class RegressionStatistics:
    def __init__(self):
        self._correct_structure_perc = None
        self._integral_err_lstsq_avg = None
        self._integral_err_om_avg = None
        self._efficiency_increase = None

    def set_correct_structure_perc(self, value):
        self._correct_structure_perc = value

    def set_integral_err_lstsq_avg(self, value):
        self._integral_err_lstsq_avg = value

    def set_integral_err_om_avg(self, value):
        self._integral_err_om_avg = value

    def set_efficiency_increase(self, value):
        self._efficiency_increase = value

    def get_correct_structure_perc(self):
        return self._correct_structure_perc

    def get_integral_err_lstsq_avg(self):
        return self._integral_err_lstsq_avg

    def get_integral_err_om_avg(self):
        return self._integral_err_om_avg

    def get_efficiency_increase(self):
        return self._efficiency_increase

class RegressionResultsAnalyzer:
    def __init__(self, regression_results):
        self._regression_results = regression_results

    def get_statistics(self) -> RegressionStatistics:
        correct_structure_count = 0
        integral_err_lstsq_sum = 0

```

```

integral_err_om_sum = 0
for result in self._regression_results:
    # print(result)
    if result.is_structure_correct():
        correct_structure_count += 1
        integral_err_lstsq_sum += result.integral_error_lstsq()
        integral_err_om_sum += result.integral_error_om()
stats = RegressionStatistics()
stats.set_correct_structure_perc(correct_structure_count / len(self._regression_results))
if correct_structure_count != 0:
    stats.set_integral_err_lstsq_avg(integral_err_lstsq_sum / correct_structure_count)
    stats.set_integral_err_om_avg(integral_err_om_sum / correct_structure_count)
    if integral_err_lstsq_sum != 0:
        stats.set_efficiency_increase((integral_err_lstsq_sum - integral_err_om_sum) / integral_err_lstsq_sum)
return stats

```

Файл original_least_squares_boost.py

```
import numpy as np
```

```
class MultivariableLinearRegressionModel:
```

```

    def __init__(self, X_data, Y_data):
        self.X_data = X_data
        self.Y_data = Y_data
        self.rows = len(X_data)
        self.cols = len(X_data[0])
        self.b_not_excluded = list(range(self.cols + 1))
        self.b_info = None

```

```

    def __add_fictive_var(self, X):
        rows = len(X)
        cols = len(X[0])
        newX = np.zeros((rows, cols+1))
        for i in range(0, rows):
            # newX[i] = np.append(X[i], (sum(X[i])/len(X[i])))
            newX[i] = np.append(X[i], -100)
        return newX

```

```

    def __remove_m2_x(self, X, M2):
        rows = len(X)
        cols = len(X[0])
        M2.sort()
        indices_set = set(M2) # Convert list to set for O(1) lookups
        modified_data = [[X[i][j]] for j in range(cols) if j not in indices_set]
            for i in range(rows)]
        return modified_data

```

```

    def __least_squares_method(self, X, Y):
        b, residuals, rank, s = np.linalg.lstsq(X, Y, rcond=None)
        return b

```

```

    def __rang_b(self, b):
        return [0] + sorted(range(1, len(b)), key=lambda x: abs(b[x]), reverse=True)

```

```

    def __split_b(self, b):
        b_rangs = self.__rang_b(b)
        M1 = [b_rangs[0], b_rangs[1]]
        M2 = [b_rangs[len(b_rangs) - 1]]
        for l in range(1, len(b_rangs) - 1):
            # average_b_m1 = sum(abs(b[b_rangs[M1[i]]]) for i in range(0, len(M1))) / (len(M1) - 1)
            average_b_m1 = sum(abs(b[b_rangs[M1[i]]]) for i in range(1, len(M1))) / (len(M1) - 1)
            if (abs(b[b_rangs[l+1]]) - abs(b[b_rangs[len(b_rangs) - 1]])) > average_b_m1 - abs(b[b_rangs[l+1]]):
                M1.append(b_rangs[l+1])
            else:
                for k in range(l+1, len(b_rangs) - 1):
                    M2.append(b_rangs[k])
        return M1, M2

```

```

    def __b_to_dict(self, matrix, column_index = 0):
        return {row_index: row[column_index] for row_index, row in enumerate(matrix)}

```

```

    def fit(self):
        newX = self.X_data
        b_info = []
        b_not_excluded = list(range(0, self.cols+1))
        while 1:
            newX = self.__add_fictive_var(newX)
            b_estimates = self.__least_squares_method(newX, self.Y_data)

```


```

b_e_dict = self.__b_to_dict(b_estimates)
if len(b_info) == 0:
    b_info = b_e_dict
    del b_info[self.cols]
M1, M2 = self.__split_b(b_estimates)
newX = self.__remove_m2_x(newX, M2)
M2.sort(reverse=True)
for el in M2:
    if el < len(b_estimates) - 1 and el < len(b_not_excluded):
        if b_not_excluded[el] < self.cols:
            b_info[b_not_excluded[el]] = 0
for el in M2:
    if el < len(b_estimates) - 1 and el < len(b_not_excluded):
        b_not_excluded.pop(el)
b_est = self.__least_squares_method(newX, self.Y_data)
if(len(M2) == 1):
    for i in range(len(b_not_excluded)):
        if b_not_excluded[i] <= self.cols - 1:
            b_info[b_not_excluded[i]] = b_est[i][0]
return b_info

```

ДОДАТОК Г

Результати перевірки роботи на співпадіння


 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute

 Дата звіту 12/8/2024
 Дата редагування 12/8/2024

☰ Документ прийнятий

Звіт подібності

метадані






Заголовок
ІП-32мп_Грицюк_ПЗ

Науковий керівник / Експерт
 Автор **Павлов О.А.**

підрозділ
ФІОТ, К-а інформатики та програмної інженерії

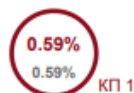
Тривога

У цьому розділі ви знайдете інформацію щодо текстових спотворень. Ці спотворення в тексті можуть говорити про **МОЖЛИВІ** маніпуляції в тексті. Спотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.

Заміна букв		0
Інтервали		0
Мікропробіли		0
Білі знаки		0
Парафрази (SmartMarks)		1

Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.



10

Довжина фрази для коефіцієнта подібності 2

15716

Кількість слів

116160

Кількість символів

Подібності за списком джерел

Нижче наведений список джерел. В цьому списку є джерела із різних баз даних. Колір тексту означає в якому джерелі він був знайдений. Ці джерела і значення Коефіцієнту Подібності не відображають прямого плагіату. Необхідно відкрити кожне джерело і проаналізувати зміст і правильність оформлення джерела.

10 найдовших фраз

Копір тексту

ПОРЯДКОВИЙ НОМЕР	НАЗВА ТА АДРЕСА ДЖЕРЕЛА URL (НАЗВА БАЗИ)	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
1	https://www.ela.kpi.ua/bitstream/123456789/53698/1/Revych_magistr.pdf	10	0.06 %
2	ІП-32мп_Галайко_ПЗ 12/2/2024 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute (ФІОТ, К-а інформатики та програмної інженерії)	9	0.06 %
3	https://www.ela.kpi.ua/bitstream/123456789/53698/1/Revych_magistr.pdf	9	0.06 %
4	https://www.ela.kpi.ua/bitstream/123456789/53698/1/Revych_magistr.pdf	8	0.05 %

Рисунок Г.1 – Результати перевірки на співпадіння (сторінка 1)




5	https://www.ela.kpi.ua/bitstream/123456789/53698/1/Revych_magistr.pdf	7	0.04 %
6	https://www.ela.kpi.ua/bitstream/123456789/53698/1/Revych_magistr.pdf	7	0.04 %
7	https://www.ela.kpi.ua/bitstream/123456789/53698/1/Revych_magistr.pdf	7	0.04 %
8	https://ela.kpi.ua/bitstream/123456789/25798/1/Linevych_magistr.pdf	7	0.04 %
9	https://ela.kpi.ua/bitstream/123456789/25798/1/Linevych_magistr.pdf	6	0.04 %
10	https://www.ela.kpi.ua/bitstream/123456789/53698/1/Revych_magistr.pdf	6	0.04 %
з домашньої бази даних (0.06 %)			
ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
1	ІП-32мн. Галайко ПЗ 12/2/2024 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute (ФІОТ, К-а інформатики та програмної інженерії)	9 (1)	0.06 %
з програми обміну базами даних (0.00 %)			
ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
з Інтернету (0.53 %)			
ПОРЯДКОВИЙ НОМЕР	ДЖЕРЕЛО URL	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
1	https://www.ela.kpi.ua/bitstream/123456789/53698/1/Revych_magistr.pdf	60 (8)	0.38 %
2	https://ela.kpi.ua/bitstream/123456789/25798/1/Linevych_magistr.pdf	13 (2)	0.08 %
3	https://docplayer.net/77616109-Ministerstvo-agrarnoyi-politiki-ukrayini-mikolajivskiy-derzhavny-agrarniy-universitet.html	5 (1)	0.03 %
4	https://ela.kpi.ua/items/8e1eef1c-9974-4410-bcc5-2aace589a0b2	5 (1)	0.03 %
Список прийнятих фрагментів (немає прийнятих фрагментів)			
ПОРЯДКОВИЙ НОМЕР	ЗМІСТ	КІЛЬКІСТЬ ОДНАКОВИХ СЛІВ (ФРАГМЕНТІВ)	

Рисунок Г.2 – Результати перевірки на співпадіння (сторінка 2)