



Теорія і методика професійної освіти

УДК 378.147

DOI <https://doi.org/10.5281/zenodo.18004752>

**Методи кореляції та регресії освітніх даних у системному аналізі якості
підготовки педагогів професійного навчання**

Деревянчук Олександр Володимирович

кандидат фізико-математичних наук, доцент, докторант кафедри професійної
підготовки, документознавства та публічного управління

Навчально-наукового інституту публічного управління та адміністрування

Українського державного університету імені Михайла Драгоманова

м. Київ, Україна

доцент кафедри професійної та технологічної освіти і загальної фізики

Чернівецького національного університету імені Юрія Федьковича

м. Чернівці, Україна

<https://orcid.org/0000-0002-3749-9998>

Прийнято: 12.10.2025 | Опубліковано: 30.10.2025

***Анотація:** У статті описано інтеграцію методів кореляції та регресії у системний аналіз якості підготовки педагогів професійного навчання, що дозволяє автоматично обробляти результати моніторингу освітнього процесу і забезпечувати зворотні зв'язки у межах логічної та фізичної моделей системного аналізу якості з метою його оптимізації.*

У дослідженні теоретично обґрунтовано доцільність впровадження методів кореляції та регресії освітніх даних у системному аналізі якості підготовки педагогів професійного навчання. Показано, що підвищити ефективність методів кореляції та регресії можливо за рахунок врахування їх



зв'язків з іншими методами інтелектуального аналізу освітніх даних (ІАОД), а також за рахунок автоматичної обробки результатів кореляційного та регресійного аналізу.

Програмну реалізацію методів кореляції та регресії, призначених для обробки освітніх даних, виконано мовою Python. Розроблена програма складається з двох модулів, які реалізують кореляційний та регресійний аналіз відповідно.

Кореляційний аналіз освітніх даних полягає в обчисленні лінійного зв'язку між двома характеристиками освітнього процесу, які математично описуються як параметри X та Y відповідно. Зв'язок між параметрами X та Y кількісно виражається коефіцієнтом кореляції Пірсона ($Corr$). Розглянуто приклад кореляційного аналізу освітніх даних, а саме семестрових оцінок здобувачів освіти із 6 предметів. Коефіцієнт кореляції $Corr$ дає змогу кількісно описати взаємозв'язки між оцінками з різних предметів і цілеспрямовано досліджувати зв'язки з високими значеннями $Corr$.

Регресійний аналіз даних полягає в обчисленні рівняння регресії, яке апроксимується поліномом степені p . Для оцінки точності апроксимації використано корінь середньої квадратичної різниці $Rmse$ для навчальної вибірки та $RmseV$ для контрольної вибірки. Оптимальну степінь полінома p обчислено як значення степеня p , для якого отримано мінімальну похибку апроксимації $RmseV$ для контрольної вибірки. Із використанням отриманого рівняння регресії виконано прогноз значень параметра Y на основі параметра X , що дозволяє, наприклад, передбачати оцінки здобувачів освіти з предмета у наступному семестрі на основі значень оцінок з предмета за попередній семестр.

Отримана модель регресії дозволяє також виконувати діагностику освітніх досягнень здобувачів освіти методом «Виявлення викидів». У випадку аналізу освітніх досягнень викид означає, що даний здобувач освіти отримав з



певного предмета оцінку значно вищу або нижчу за прогнозовану і навчальні результати такого здобувача освіти потребують подальших досліджень.

Показано, що за допомогою розробленої програми можливо проводити кореляційний та регресійний аналіз, а також прогнозування, виявлення взаємозв'язків і викидів освітніх даних для вибірки великого розміру (1000 здобувачів освіти). Впровадження методів кореляції та регресії дає змогу автоматичного обробляти результати моніторингу освітнього процесу з метою забезпечення якості підготовки педагогів професійного навчання.

Ключові слова: *системний аналіз, якість освіти, педагоги професійного навчання, методи кореляції, методи регресії, прогнозування, інтелектуальний аналіз освітніх даних.*

Correlation and regression methods for educational data in the systems analysis of the quality of training of vocational education teachers

Oleksandr Derevyanchuk

Candidate of Physical and Mathematical Sciences, Associate Professor,
Doctoral Candidate of the Department of Professional Training, Document Science,
and Public Administration Educational and Scientific Institute of Public
Administration and Management Dragomanov Ukrainian State University
Kyiv, Ukraine

Associate Professor of the Department Professional and Technological Education and
General Physics, Yuriy Fedkovych Chernivtsi National University
Chernivtsi, Ukraine

<https://orcid.org/0000-0002-3749-9998>

Abstract: *The article describes the integration of correlation and regression methods into the systems analysis of the quality of training of vocational education*



teachers, which enables automated processing of educational process monitoring results and provides feedback loops within the logical and physical models of the systems analysis of quality for the purpose of its optimization.

The study theoretically substantiates the expediency of implementing correlation and regression methods for educational data in the systemic analysis of the quality of vocational education teachers' training. It is shown that the effectiveness of correlation and regression methods can be increased by taking into account their relationships with other methods of Educational Data Mining (EDM), as well as through automatic processing of the results of correlation and regression analysis.

The software implementation of correlation and regression methods intended for processing educational data was carried out in Python. The developed program consists of two modules that implement correlation analysis and regression analysis, respectively.

Correlation analysis of educational data involves computing the linear relationship between two characteristics of the educational process, which are mathematically described as parameters X and Y , respectively. The relationship between parameters X and Y is quantitatively expressed by Pearson's correlation coefficient ($Corr$). An example of correlation analysis of educational data is considered, namely semester grades of students in six subjects. The correlation coefficient $Corr$ makes it possible to quantify the relationships between grades in different subjects and to purposefully investigate relationships with high $Corr$ values.

Regression analysis involves computing a regression equation that is approximated by a polynomial of degree p . To assess approximation accuracy, the root mean square error $Rmse$ was used for the training sample and $RmseV$ for the validation sample. The optimal polynomial degree pA was determined as the value of p for which the minimum approximation error $RmseV$ was obtained on the validation sample. Using the resulting regression equation, the values of parameter Y were forecast based



on parameter X , which makes it possible, for example, to predict students' grades in a subject for the next semester based on their grades in the previous semester.

The obtained regression model also enables the diagnosis of students' educational achievements using the outlier detection method. In the context of analyzing educational achievements, an outlier means that a given student received a grade in a particular subject that is significantly higher or lower than the predicted one, and the learning outcomes of such a student require further investigation.

It is shown that the developed program makes it possible to perform correlation and regression analysis, as well as forecasting, identification of relationships, and detection of outliers in educational data for a large sample (1,000 students). Implementing correlation and regression methods enables automated processing of educational process monitoring results in order to ensure the quality of training of vocational education teachers.

Keywords: *systems analysis, quality of education, vocational education teachers, correlation methods, regression methods, forecasting, educational data mining.*

Постановка проблеми. Якість підготовки педагогів професійного навчання залежить від множини факторів, зокрема, від методичного та матеріально-технічного забезпечення освітнього процесу, а також від навчальних досягнень здобувачів освіти. Забезпечення високого рівня якості підготовки педагогів професійного навчання потребує як врахування вимог ринку праці, які динамічно змінюються, так і впровадження сучасних цифрових технологій.

На даний час, за рахунок застосування цифрових технологій, обсяг освітніх даних постійно зростає, що потребує застосування автоматизованих методів аналізу таких даних, а саме методів інтелектуального аналізу освітніх даних (ІАОД) [1]. Методи ІАОД є підмножиною методів інтелектуального аналізу



даних (ІАД), які адаптовані до особливостей освітнього процесу.

До важливих і поширених методів ІАОД належать, зокрема, методи прогнозування, кластеризації, класифікації, кореляційного та регресійного аналізу [2, 3]. Практична реалізація таких методів виконується за допомогою відповідних апаратно-програмних засобів та технологій. Проте, окреме застосування методів ІАОД має обмежену ефективність для забезпечення якості підготовки педагогів професійного навчання, оскільки не враховує всіх основних зв'язків між характеристиками освітнього процесу та не забезпечує цілісної й комплексної підготовки здобувачів освіти.

Ефективність підготовки педагогів професійного навчання можливо підвищити за рахунок системного аналізу якості їх підготовки, який передбачає комплексне, узгоджене та цілеспрямоване застосування методів, засобів і технологій ІАОД. Таким чином, можливо підвищити ефективність методів кореляції та регресії, які широко застосовуються при обробці освітніх даних, за рахунок врахування зв'язків даних методів з іншими методами ІАОД, а також за рахунок автоматичного вибору параметрів методів кореляції та регресії при їх прикладному застосуванні.

Завдання даного дослідження, яке полягає в інтеграції методів кореляції та регресії у системний аналіз якості підготовки педагогів професійного навчання, є актуальним. Застосування методів кореляції та регресії дає змогу автоматичного обробляти результати моніторингу освітнього процесу, забезпечувати зворотні зв'язки у межах логічної та фізичної моделей системного аналізу якості підготовки педагогів професійного навчання. Автоматичний вибір параметрів для методів кореляції та регресії підвищує швидкодю та надійність їх практичного застосування. Метод кореляції може успішно застосовуватися для аналізу взаємозв'язків між даними. Урахування зв'язків методу регресії з методами прогнозування та аналізу викидів розширює можливості системного аналізу якості підготовки педагогів професійного навчання.



Аналіз останніх публікацій. Огляд сучасної літератури засвідчує зростання інтересу до впровадження методів кореляції та регресії в освітній процес, що є важливим у контексті системного аналізу якості підготовки педагогів професійного навчання.

У роботі М.З. Згуровського, Н.Д. Панкратова [4] зазначено, що однією з процедур системного аналізу є структурно-функціональний аналіз, який «дає змогу визначити необхідний рівень потенційних можливостей функціональних елементів складної системи і ступінь взаємозв'язків і взаємозалежностей її функціональних елементів для досягнення заданих цілей функціонування системи в ситуації, що складається апріорі» [2]. Тому для урахування та кількісного опису таких взаємозв'язків і взаємозалежностей між елементами логічної та фізичної моделей системного аналізу якості є доцільним застосування методів кореляції та регресії.

У дослідженні П. М. Грицюка, О. І. Джоші, О. М. Гладкої [5] системний аналіз розглядається як наукова дисципліна, призначенням якої є практичне застосування результатів теорії систем до задач керування складними системами, зокрема, освітніми. У системному аналізі метод конкретизується як засіб, що у випадку застосування методів кореляції та регресії для системного аналізу якості підготовки педагогів професійного навчання означає їх реалізацію у вигляді прикладних програм.

Регресійний аналіз даних та виявлення структурних взаємозв'язків між змінними описано у роботі Г. Барлетта, Дж. Трецца, Е. К'яваццо (*G. Barletta, G. Trezza ma E. Chiavazzo*) [6]. У контексті системного аналізу якості освіти ця робота підтверджує доцільність використання методу регресії для кількісного опису взаємозв'язків між освітніми параметрами, наприклад, при побудові фізичної моделі освітнього аналізу якості.

У роботі [7, 8] описано основи кореляційного та регресійного аналізу. Регресійний аналіз розглянуто як розділ математичної статистики, присвячений



методам аналізу залежності математичного очікування однієї величини від значень іншої. Показано, що на відміну від кореляційного аналізу, регресійний аналіз не з'ясовує чи істотний зв'язок, а займається пошуком моделі цього зв'язку, вираженої у функції регресії.

Застосування регресійного аналізу для прогнозування описано у роботах [9, 10], в яких залежність прогнозованих значень (залежні змінні) між початкових параметрів (незалежні змінні) математично описано рівнянням регресії (поліномом). Коефіцієнти рівняння регресії обчислено методом найменших квадратів. Прогнозування на прикладі розміщення студентів коледжу на основі академічної успішності за регресійного аналізу описано в роботі [11].

У сучасних системах ІАОД, крім регресійного аналізу, для прогнозування часто застосовуються штучні нейронні мережі (ШНМ) [12-14]. Використання штучних нейронних мереж для прогнозування результатів освітнього процесу є ефективним у випадку складних залежностей між параметрами. Проте, перед застосуванням ШНМ потребують навчання на основі спеціально підібраних наборів даних.

Проведений огляд наукових публікацій підтверджує, що впровадження методів кореляції та регресії у системний аналіз якості підготовки педагогів професійного навчання дає змогу кількісно описати взаємозв'язки між характеристиками освітнього процесу. При реалізації концептуальної, логічної та фізичної моделей системного аналізу якості, методи кореляції та регресії дозволяють автоматично обробляти результати моніторингу освітнього процесу і забезпечувати зворотні зв'язки, завдяки чому, в цілому, підвищується якість освітнього процесу на всіх рівнях (від освітніх програм до здобувачів освіти) [15, 16]. Це сприяє підвищенню обґрунтованості управлінських рішень, впровадженню індивідуальних освітніх траєкторій, реалізації прогнозування та виявлення способів підвищення освітніх досягнень здобувачів освіти.

Виділення невирішених раніше частин загальної проблеми.



Незважаючи на наявність досліджень за тематикою впровадження методів кореляції та регресії в освітній процес, недостатньо опрацьованими є питання прикладного застосування таких методів у системному аналізі якості підготовки педагогів професійного навчання. Потребує подальших досліджень питання автоматичного вибору параметрів регресійного аналізу, застосування кореляційного аналізу для встановлення взаємозв'язків між освітніми даними, комплексного застосування результатів кореляційного та регресійного аналізу, застосування методу регресії для прогнозування та аналізу викидів освітніх даних.

Мета статті полягає в інтеграції методів кореляції та регресії у системний аналіз якості підготовки педагогів професійного навчання, що дозволяє автоматично обробляти результати моніторингу освітнього процесу і забезпечувати зворотні зв'язки в межах логічної та фізичної моделей системного аналізу якості з метою його оптимізації.

Виклад основного матеріалу. Розроблено програму «regres_ed_25», яка призначена для кореляційного та регресійного аналізу освітніх даних, а також встановлення взаємозв'язків, прогнозування та виявлення викидів. Програму реалізовано мовою Python. Програма призначена для впровадження у системний аналіз якості підготовки педагогів професійного навчання. За структурою програма складається з двох модулів, які реалізують кореляційний та регресійний аналіз відповідно.

1. Програмна реалізація кореляційного аналізу освітніх даних. Кореляційний аналіз освітніх даних полягає в обчисленні лінійного зв'язку (поліномом степеня $p = 1$) між двома характеристиками освітнього процесу, які математично описуються як параметри X та Y відповідно. Як параметри освітнього процесу розглядаються, наприклад, проміжні та підсумкові оцінки здобувачів освіти з певних навчальних предметів, результати тестування та опитування. Зв'язок між параметрами X та Y кількісно виражається коефіцієнтом

кореляції Пірсона (*Corr*) [16, 17], який обчислюється для значень навчальної вибірки за формулою:

$$Corr = \frac{\sum_{i=0}^{N-1} (x_i - X_s)(y_i - Y_s)}{\sqrt{\sum_{i=0}^{N-1} (x_i - X_s)^2} \cdot \sqrt{\sum_{i=0}^{N-1} (y_i - Y_s)^2}}. \quad (1)$$

де x_i, y_i – значення з номером i для параметрів X та Y відповідно;

N – кількість значень параметрів;

X_s, Y_s – середні значення параметрів X та Y .

Розглянемо приклад обробки освітніх даних розробленої програмою. Шляхом кореляційного аналізу обчислено лінійний зв'язок між оцінками здобувачів освіти з різних предметів на 1-й семестр (параметр X) та їх оцінками за 3-й семестр (параметр Y). Початкові файли для аналізу зчитуються з таблиць, в яких у кожному стовпчику записані значення певного досліджуваного параметра. Таблиці з даними зберігаються як текстові файли або як книги Microsoft Excel. Наприклад, семестрові оцінки здобувачів освіти з навчальних предметів зчитано з файлу Microsoft Excel «Оцінки_24.xlsx» (табл. 1).

Таблиця 1

Фрагмент файлу «Оцінки_24. Xlsx»

Інженерна графіка	Вища математика	Інформатика	Загальна електротехніка	Іноземна мова	Психологія
65	60	70	80	50	60
90	74	60	88	71	70
90	90	90	75	70	65
90	80	63	90	77	65
92	64	70	80	50	60
62	72	54	65	51	79
90	80	90	92	81	94
90	82	70	90	85	96
92	92	90	95	93	100
65	60	65	65	50	55

Джерело: розробка автора на основі використання семестрових оцінок здобувачів освіти із 6-ти предметів

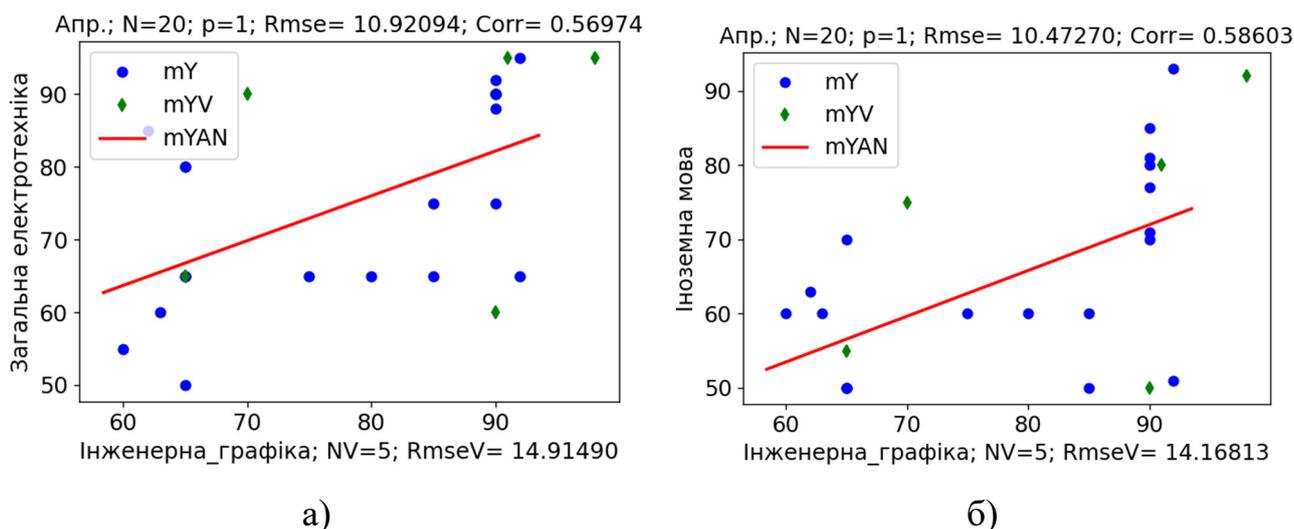
У файлі «Оцінки_24. Xlsx» збережено семестрові оцінки здобувачів освіти із 6 предметів:

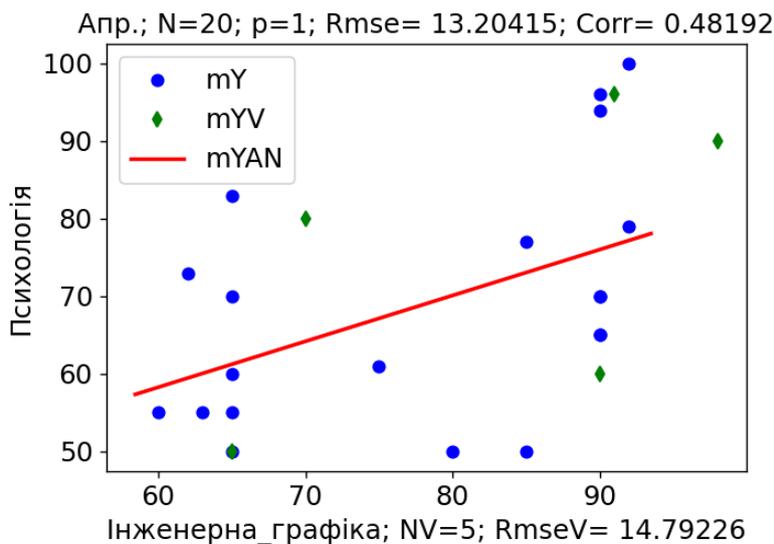
1. Інженерна графіка.
2. Вища математика.
3. Інформатика.
4. Загальна електротехніка.
5. Іноземна мова.
6. Психологія.

Перших три предмети викладалися на 1-му курсі (1-й семестр), а наступних три предмети викладалися на 2-му курсі (3-й семестр). Усі досліджувані дані розділено на навчальну (N значень) та контрольну (NV значень) вибірки. Навчальна вибірка призначена для кореляційного та регресійного аналізу даних, а контрольна вибірка – для перевірки результатів такого аналізу. Значення параметрів навчальної вибірки записувалися в масиви mX та mY , а значення параметрів контрольної вибірки записувалися в масиви mXV та mYV . Отримано такі результати кореляційного аналізу даних (рис. 1 – 3).

Рисунок 1

Лінійна апроксимація початкових даних та значення їх коефіцієнта кореляції $Corr$, досліджено зв'язок між оцінками з предмета «Інженерна графіка» (1-й семестр) та оцінками з предметів за 3-й семестр



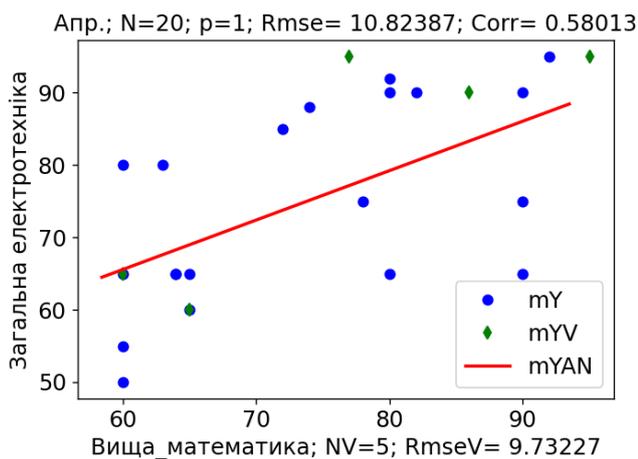


в)

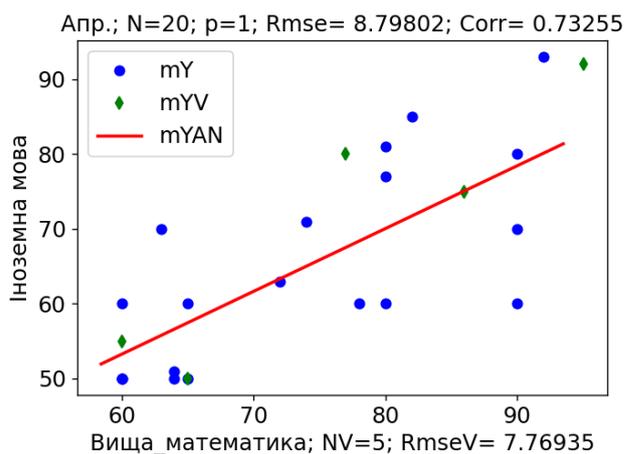
Джерело: розробка автора на основі використання програми «regres_ed_25»

Рисунок 2

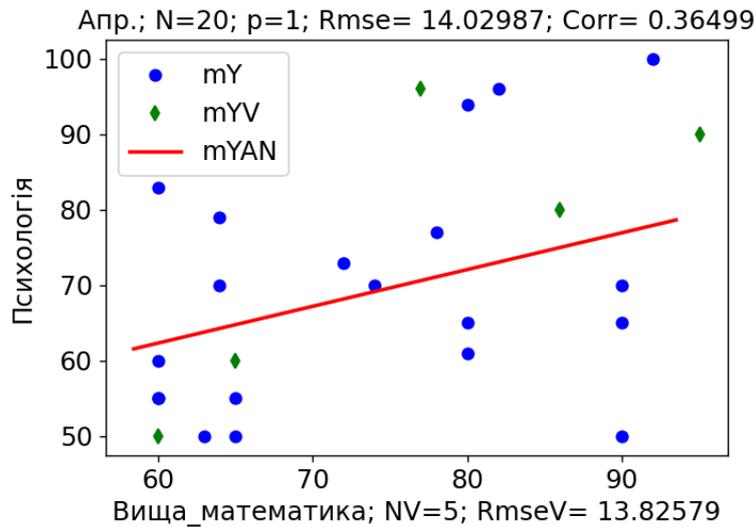
Лінійна апроксимація початкових даних та значення їх коефіцієнта кореляції Corr, досліджено зв'язок між оцінками з предмета «Вища математика» (1-й семестр) та оцінками з предметів за 3-й семестр



а)



б)

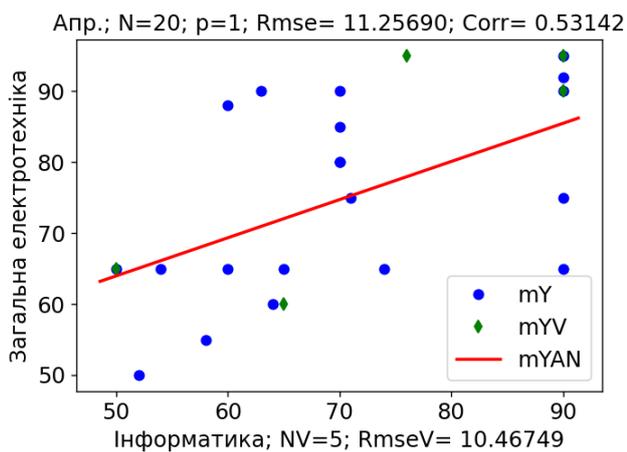


в)

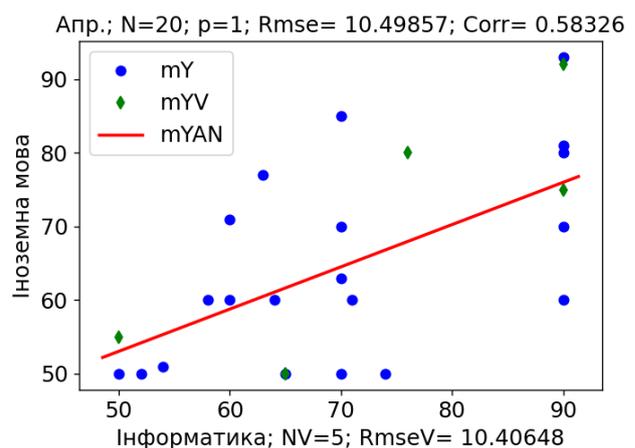
Джерело: розробка автора на основі використання програми «regres_ed_25»

Рисунок 3

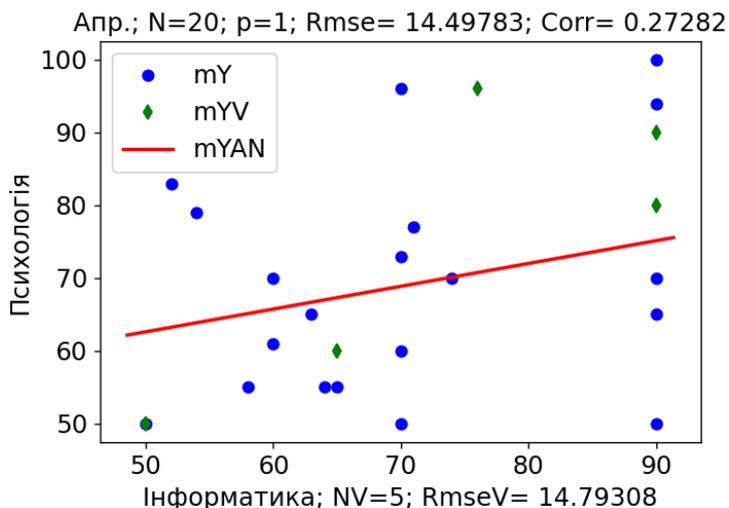
Лінійна апроксимація початкових даних та значення їх коефіцієнта кореляції Corr, досліджено зв'язок між оцінками з предмета «Інформатика» (1 -й семестр) та оцінками з предметів за 3-й семестр



а)



б)



в)

Джерело: розробка автора на основі використання програми «regres_ed_25»

Значення коефіцієнта кореляції *Corr* завжди знаходяться в діапазоні від -1 до +1. Значення, близькі до +1, вказують на сильну позитивну кореляцію (зі зростанням одного параметру другий також зростає), а значення, близькі до -1, вказують на сильну негативну кореляцію (зі зростанням одного параметру інший зменшується). Значення, близьке до 0, свідчить про відсутність лінійного зв'язку.

Підсумкові результати кореляційного аналізу оцінок з навчальних предметів (на основі рис. 1 – 3) наведено в табл. 2 та на рис. 4.

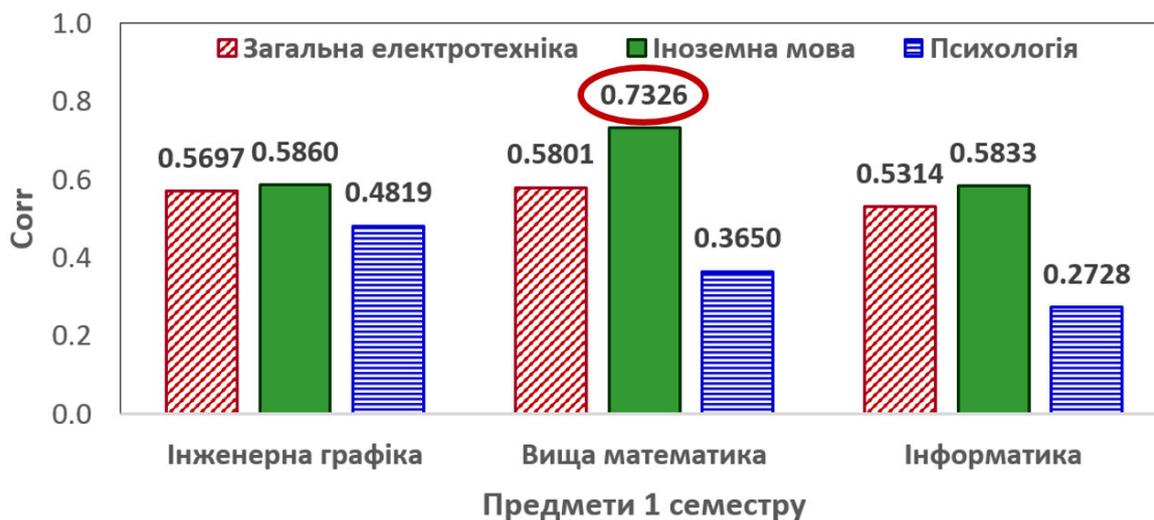
Таблиця 2.

Значення коефіцієнтів кореляції Corr для оцінок з предметів

Предмети 1-ого семестру \ Предмети 3-ого семестру	Загальна електротехніка	Іноземна мова	Психологія
Інженерна графіка	0.5697	0.5860	0.4819
Вища математика	0.5801	0.7326	0.3650
Інформатика	0.5314	0.5833	0.2728

Рисунок 4

Значення коефіцієнтів кореляції *Corr* для оцінок з предметів



Джерело: розробка автора на основі використання програми «regres_ed_25»

Коефіцієнт кореляції *Corr* дає змогу кількісно описати взаємозв'язки між оцінками з різних предметів і цілеспрямовано досліджувати зв'язки з високими значеннями *Corr* (у деяких випадках високі значення *Corr* можуть означати причинно-наслідкові зв'язки між параметрами).

Найвище значення коефіцієнта кореляції *Corr* отримано для оцінок з предметів «Вища математика» та «Іноземна мова» (рис. 2, б; рис. 4), тому в подальшому проведено регресійний аналіз взаємозв'язку освітніх результатів саме з цих предметів. Високе значення коефіцієнта кореляції *Corr* для оцінок з предметів «Вища математика» та «Іноземна мова» можна пояснити тим, що високі результати з математики (1-й семестр) отримують здобувачі освіти з розвинутим логічним мисленням, здібностями до запам'ятовування, наполегливістю та дисциплінованістю. Такі характеристики дозволяють здобувачам освіти отримувати високі оцінки також з іноземної мови (3-й семестр).

2. Програмна реалізація регресійного аналізу освітніх даних. Рівняння регресії, яке описує зв'язок параметрів $Y(X)$, побудовано у вигляді полінома [7]. Тому досліджувану залежність $Y(X)$ апроксимовано поліномом степені p

$$y_A(x) = a_0 + \sum_{n=1}^p a_n \cdot x^n. \quad (2)$$

Апроксимовані значення y_A записано в масиви $mYAN$ (рис. 1 – 3). Коефіцієнти полінома $a_0...a_p$ обчислено методом найменших квадратів. Для оцінки точності апроксимації використано корінь середньої квадратичної різниці $Rmse$ (Root mean square error) для навчальної вибірки та $RmseV$ для контрольної вибірки, які обчислювалися за формулами

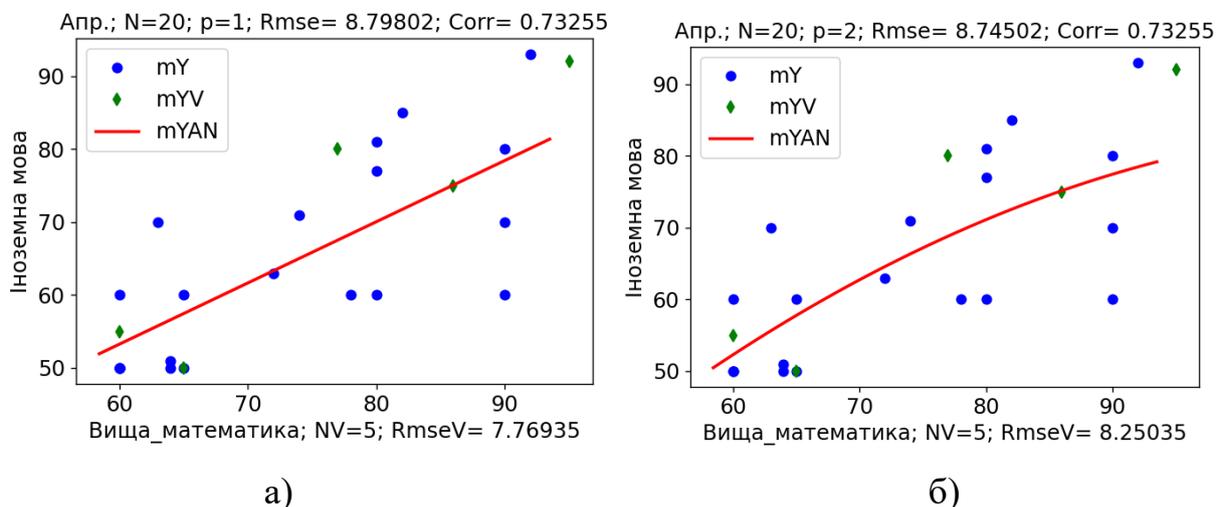
$$Rmse = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (y(i) - y_A(i))^2}. \quad (3)$$

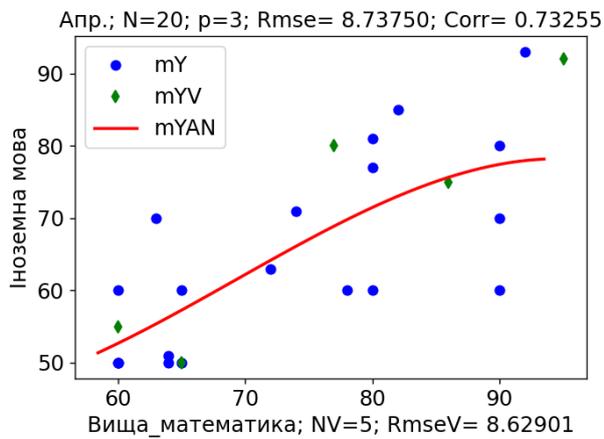
$$RmseV = \sqrt{\frac{1}{N_V} \sum_{i=0}^{N_V-1} (y_V(i) - y_{AV}(i))^2}. \quad (4)$$

Автоматичне визначення оптимальної степені полінома для рівняння регресії виконано шляхом перебору значень у заданому діапазоні від 1 до 6 (рис. 5).

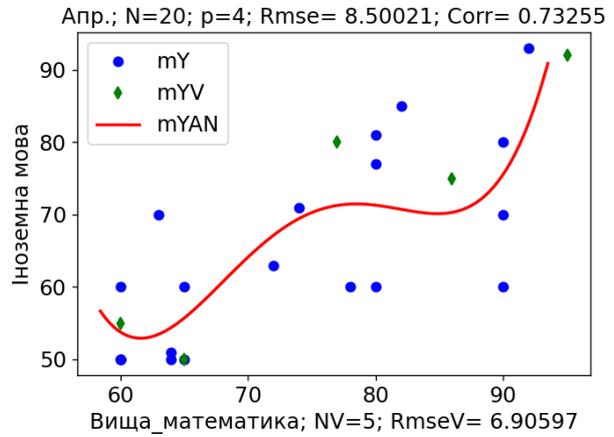
Рисунок 5

Регресійний аналіз даних поліномом степеня p , який змінювався у діапазоні від 1 до 6

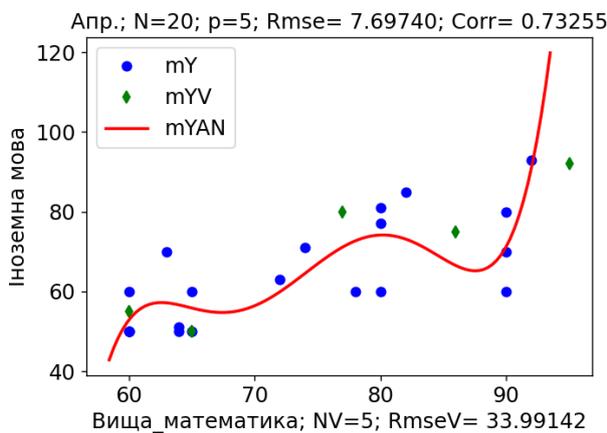




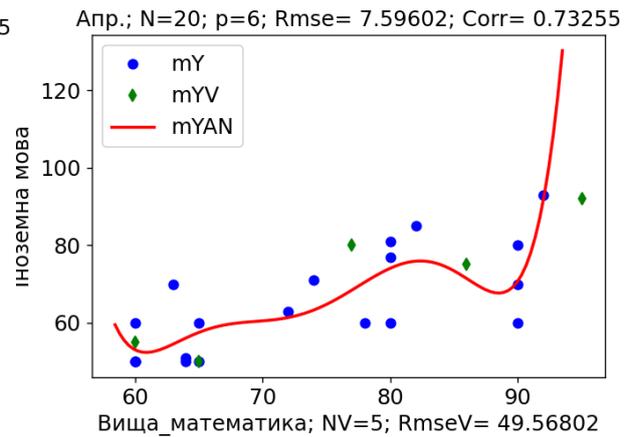
В)



Г)



Д)



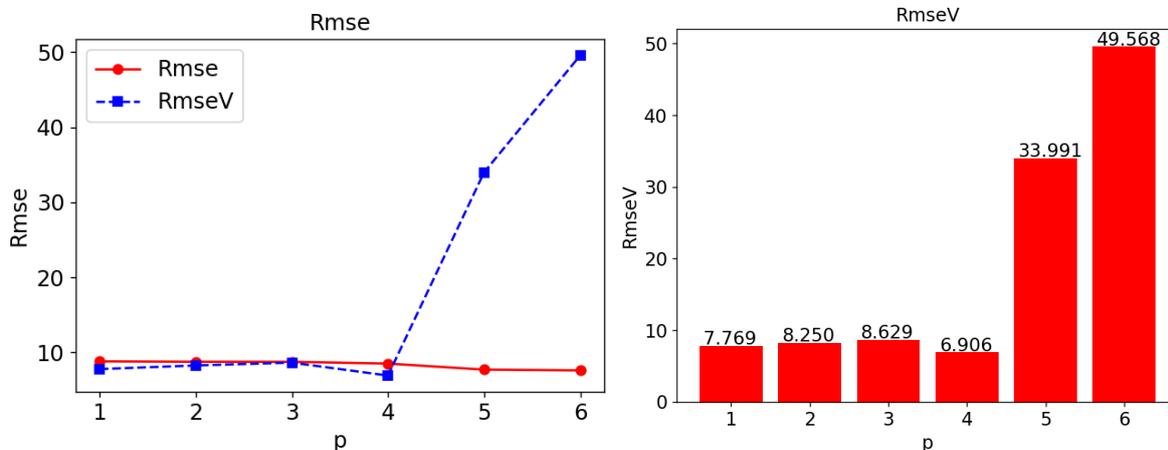
Е)

Джерело: розробка автора на основі використання програми «regres_ed_25»

Оптимальний степінь полінома pA обчислений як значення p , для якого отримано мінімальну похибку апроксимації $RmseV$ для контрольної вибірки (рис. 6).

Рисунок 6

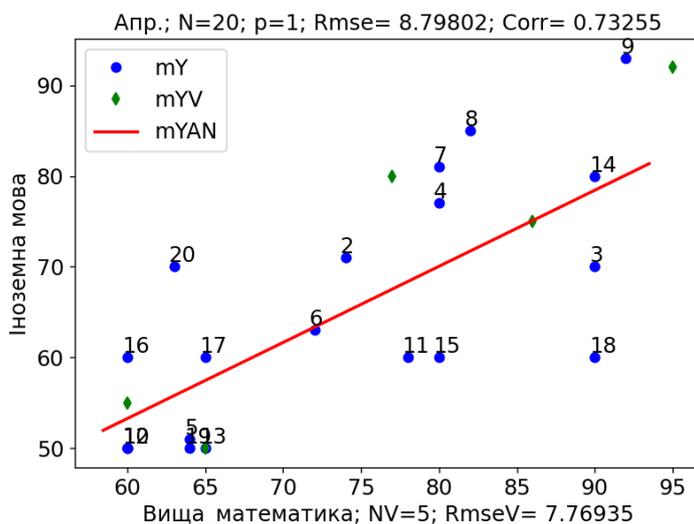
Залежність похибок апроксимації для навчальної вибірки ($Rmse$) та контрольної вибірки ($RmseV$) залежно від степеня полінома p



Джерело: розробка автора на основі використання програми «regres_ed_25»

Рисунок 7

Регресійний аналіз початкових даних для оптимального степеня полінома $pA = 1$ із указанням номерів точок



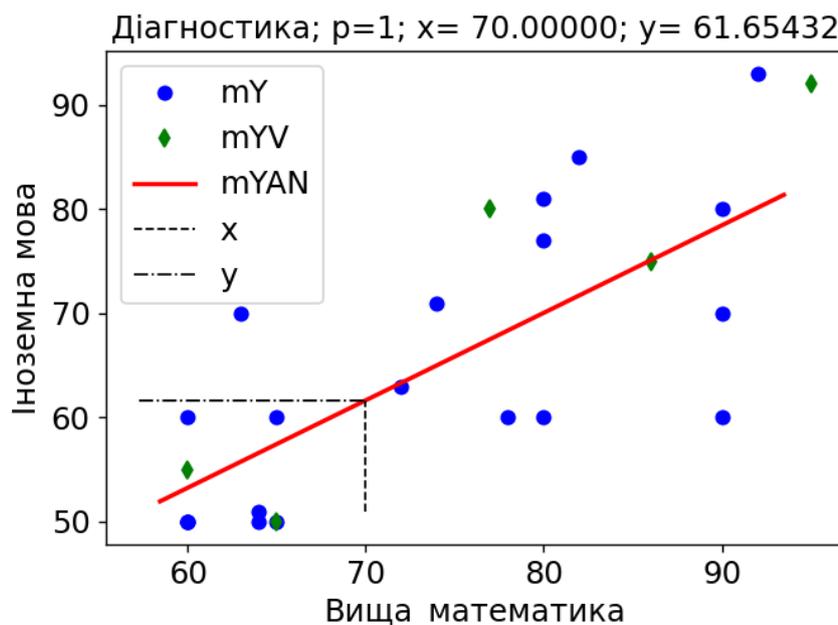
Джерело: розробка автора на основі використання програми «regres_ed_25»

Крім цього, для отриманого полінома перевірялася умова монотонності (для того, щоб встановити однозначні залежності між досліджуваними величинами). Мінімальне значення $RmseV$ отримано при $p = 4$ (рис. 5, г), проте для такого степеня поліном отримується не монотонним. Тому, як оптимальне значення степеня полінома отримано $pA = 1$ (рис. 5, а; рис. 7).

Із використанням отриманого рівняння регресії (2), виконано прогноз значень параметра Y на основі параметра X (рис. 8). Практичне застосування такого прогнозу дозволяє з певною точністю передбачати оцінки здобувачів освіти у наступному семестрі (наприклад, з предмету «Іноземна мова») на основі значень оцінок за попередній семестр (наприклад, з предмета «Вища математика»). У розглянутому випадку (рис. 8) здобувачу освіти, який отримав у 1-му семестрі оцінку 70 балів з предмета «Вища математика», у 3-му семестрі з предмета «Іноземна мова» прогнозується оцінка 62 бали.

Рисунок 8

Прогноз значень параметра Y (Іноземна мова) на основі параметра X (Вища математика)



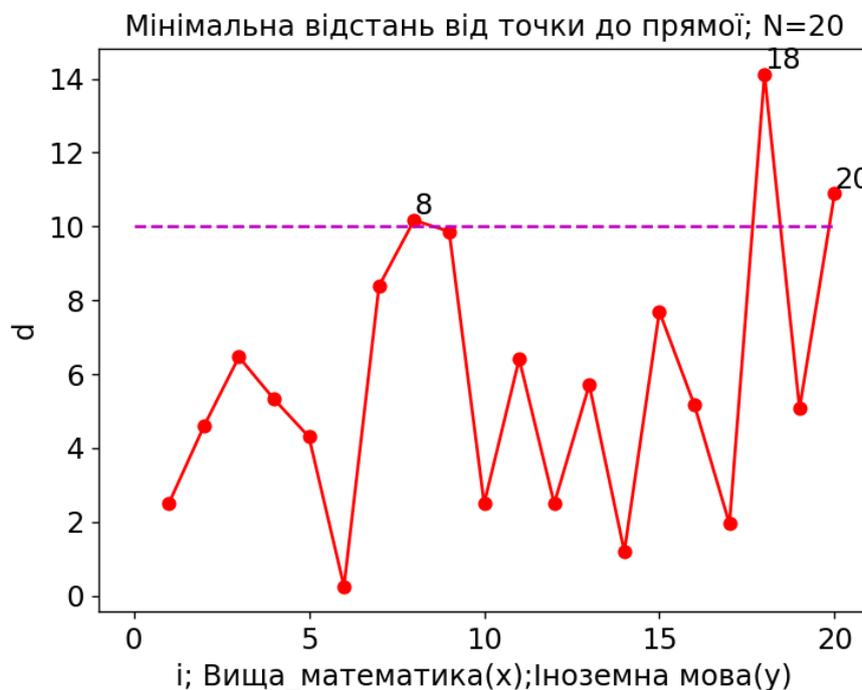
Джерело: розробка автора на основі використання програми «regres_ed_25»

Отримана модель регресії дозволяє також виконувати діагностику освітніх досягнень здобувачів освіти методом «Виявлення викидів» (Outlier Detection), який належить до методів ІАОД. З цією метою, для кожного досліджуваного об'єкта, обчислено модуль мінімальної відстані d точки об'єкта до рівняння регресії (рис. 9).

Об'єкт класифікується як викид, якщо відстань перевищує встановлений поріг Th_d . У випадку аналізу освітніх досягнень викид означає, що даний здобувач освіти отримав з певного предмета оцінку значно вищу або нижчу за прогнозовану. Тому навчальні результати таких здобувачів освіти потребують подальших досліджень з метою підвищення якості освітнього процесу.

Рисунок 9

Значення модуля мінімальної відстані d точки від рівняння регресії; поріг $Th_d = 10$

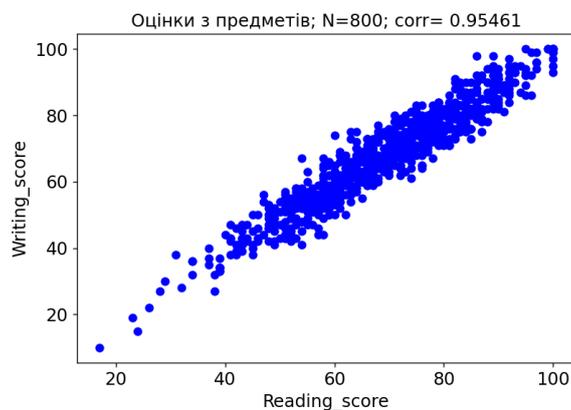


Джерело: розробка автора на основі використання програми «regres_ed_25»

3. Кореляційний та регресійний аналіз освітніх даних великого розміру. За допомогою розробленої програми «regres_ed_25» проведено кореляційний та регресійний аналіз, а також виявлення викидів для вибірки з 1000 здобувачів освіти, де навчальна вибірка містила 800 здобувачів освіти, а контрольна вибірка – 200 здобувачів освіти [19]. Початкові дані отримано з файлу «Reading_Writing_Score.txt» і досліджено зв'язок оцінок з письма «Writing_score» (параметр Y) та оцінок з читання «Reading_score» (параметр X) (рис. 10). Кореляційний аналіз показує сильний зв'язок параметрів, про що свідчить високе значення коефіцієнту кореляції $Corr$.

Рисунок 10

Значення оцінок з письма «Writing_score» (Y) та оцінок з читання «Reading_score» (X) для навчальної вибірки здобувачів освіти

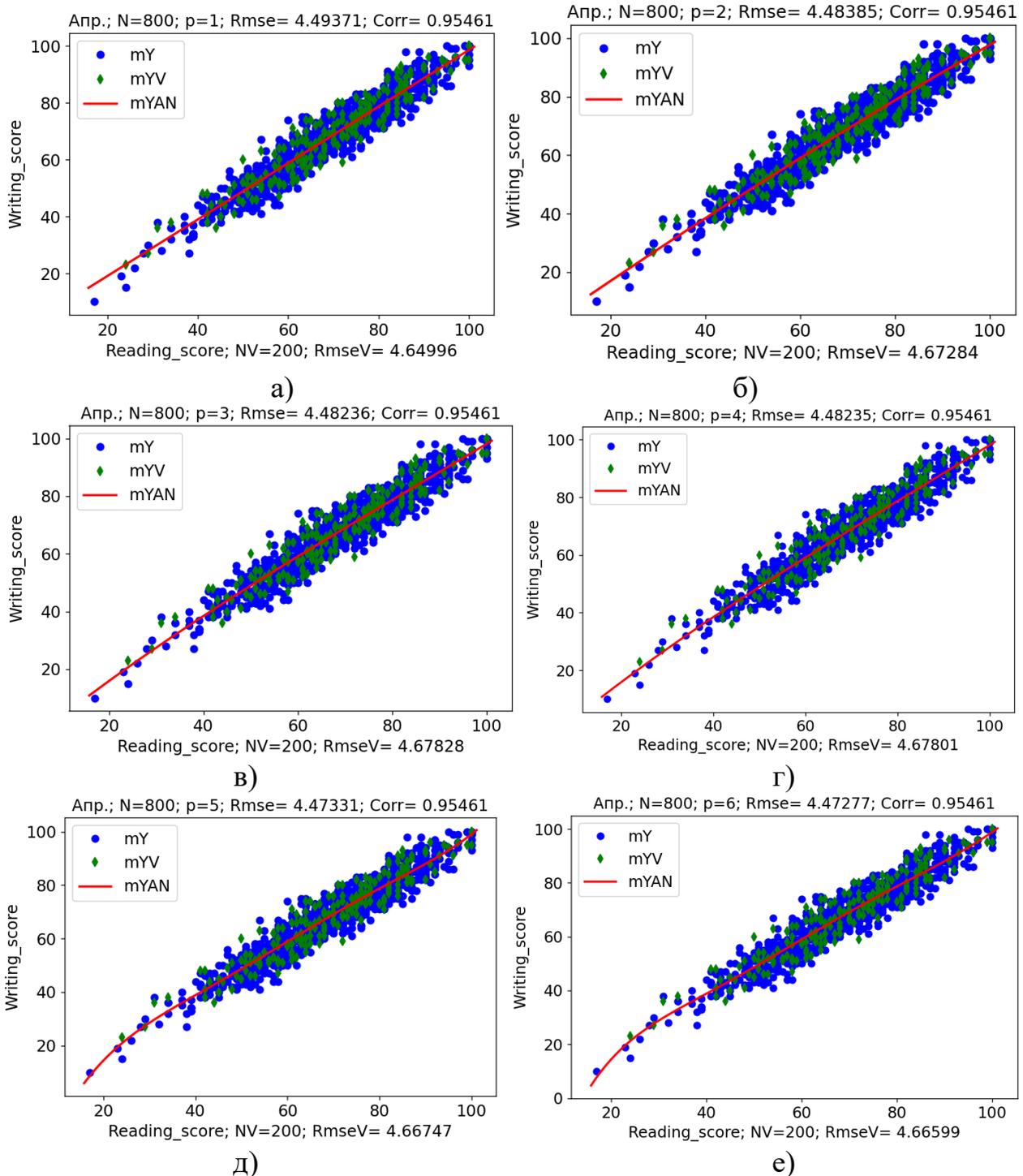


Джерело: розробка автора на основі використання програми «regres_ed_25»

Автоматичне визначення оптимального степеня полінома для рівняння регресії виконано шляхом перебору значень у заданому діапазоні від 1 до 6 (рис. 11).

Рисунок 11

Регресійний аналіз даних (рис. 10) поліномом степеня p , який змінювався у діапазоні від 1 до 6



Джерело: розробка автора на основі використання програми

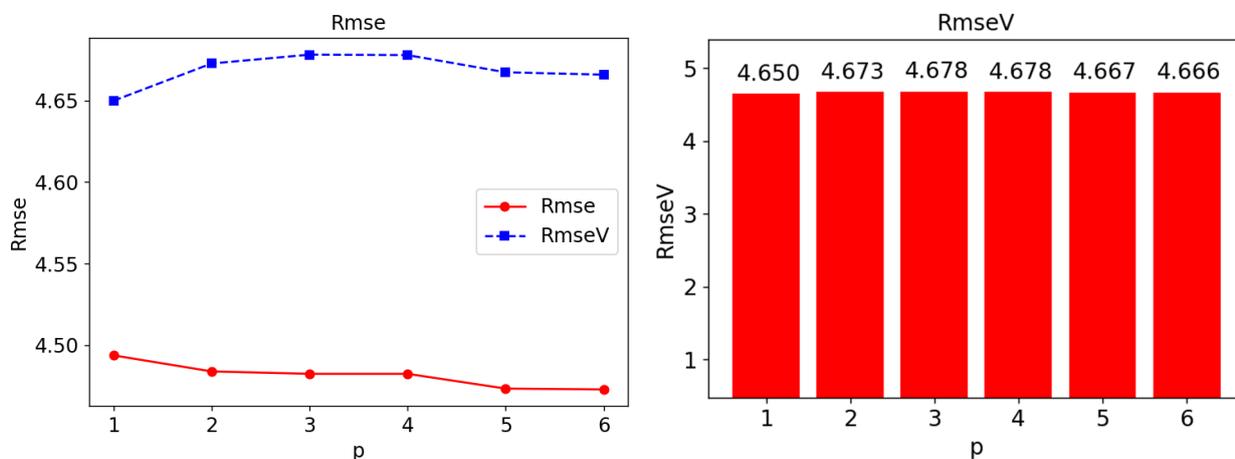
«regres_ed_25»

За мінімальним значенням похибки апроксимації $RmseV$ (рис. 12) отримано оптимальне значення степеня полінома $pA = 1$ (рис. 11, а).

Із використанням рівняння регресії (рис. 11, а) виконано прогноз значень параметра Y на основі параметра X (рис. 13). У розглянутому випадку, здобувачу освіти, який отримав оцінку 70 балів з читання «Reading_score», з письма «Writing_score» прогнозується оцінка 69 балів.

Рисунок 12

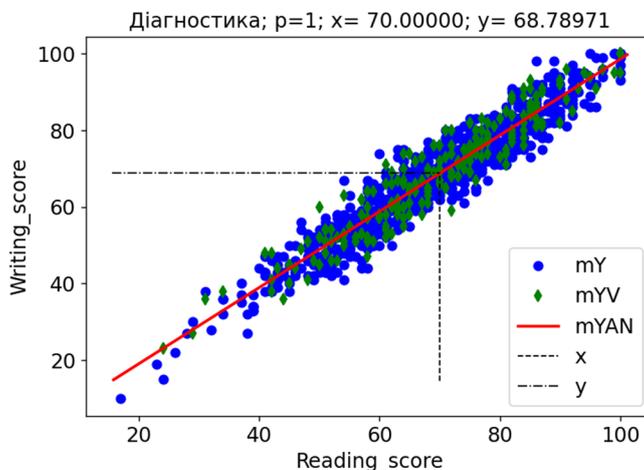
Залежність похибок апроксимації для навчальної вибірки ($Rmse$) та контрольної вибірки ($RmseV$) залежно від степеня полінома p



Джерело: розробка автора на основі використання програми «regres_ed_25»

Рисунок 13

Прогноз значень параметра Y «Writing_score» на основі параметра X «Reading_score»

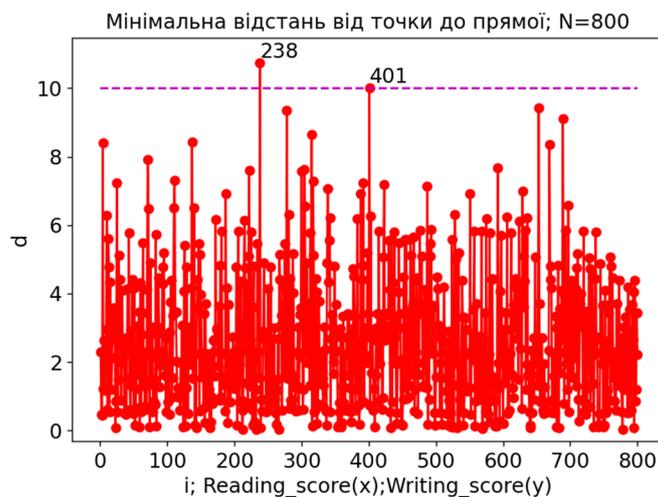


Джерело: розробка автора на основі використання програми «regres_ed_25»

Отримана модель регресії дозволяє також виявляти викиди даних (рис. 14). З цією метою для кожного досліджуваного об'єкта обчислено модуль мінімальної відстані d точки об'єкта до рівняння регресії. Об'єкт класифікується як викид, якщо модуль мінімальної відстані d точки об'єкта до рівняння регресії перевищує встановлений поріг Th_d . У даному випадку викид означає, що здобувач освіти отримав з певного предмета оцінку значно вищу або нижчу за прогнозовану.

Рисунок 14

Значення модуля мінімальної відстані d точки від рівняння регресії; поріг $Th_d = 10$



Джерело: розробка автора на основі використання програми «regres_ed_25»

Таким чином, розроблена програма «regres_ed_25» дає змогу ефективно виконувати кореляційний та регресійний аналіз, а також прогнозування, виявлення взаємозв'язків і викидів освітніх даних. Програма дозволяє обробляти вибірки значного розміру (до 10^6 елементів), що є достатнім для практичного застосування. У програмі застосовується зручна візуалізація даних із застосуванням методів інфографіки. Поділ даних на навчальну та контрольну



вибірки дозволяє об'єктивно визначати параметри для регресійного аналізу даних. Автоматичний вибір степеня полінома для рівняння регресії дає змогу підвищити швидкодію регресійного аналізу даних.

Висновки. У дослідженні теоретично обґрунтовано доцільність впровадження методів кореляції та регресії освітніх даних у системному аналізі якості підготовки педагогів професійного навчання. Показано, що підвищити ефективність методів кореляції та регресії можливо за рахунок врахування зв'язків даних методів з іншими методами ІАОД, а також за рахунок автоматичного вибору параметрів методів кореляції та регресії при їх прикладному застосуванні.

Програмну реалізацію методів кореляційного та регресійного аналізу освітніх даних виконано мовою Python. Розроблена програма складається з двох модулів, які реалізують кореляційний та регресійний аналіз відповідно.

Кореляційний аналіз освітніх даних полягає в обчисленні лінійного зв'язку між двома характеристиками освітнього процесу, які математично описуються як параметри X та Y відповідно. Зв'язок між параметрами X та Y кількісно виражається коефіцієнтом кореляції Пірсона ($Corr$). Розглянуто приклад кореляційного аналізу освітніх даних, а саме семестрових оцінок здобувачів освіти з 6-ти предметів. Коефіцієнт кореляції $Corr$ дає змогу кількісно описати взаємозв'язки між оцінками з різних предметів і цілеспрямовано досліджувати зв'язки з високими значеннями $Corr$.

Регресійний аналіз даних полягає в обчисленні рівняння регресії, яке апроксимується поліномом степеня p . Для оцінки точності апроксимації використано корінь середньої квадратичної різниці $Rmse$ для навчальної вибірки та $RmseV$ для контрольної вибірки. Оптимальну степінь полінома pA обчислено як значення степеня p , для якого отримано мінімальну похибку апроксимації $RmseV$ для контрольної вибірки. Із використанням отриманого рівняння регресії виконано прогноз значень параметра Y на основі параметра X , що дозволяє,



наприклад, передбачати оцінки здобувачів освіти з предмета у наступному семестрі на основі значень оцінок з предмета за попередній семестр.

Отримана модель регресії дозволяє також виконувати діагностику освітніх досягнень здобувачів освіти методом «Виявлення викидів». У випадку аналізу освітніх досягнень викид означає, що даний здобувач освіти отримав з певного предмета оцінку значно вищу або нижчу за прогнозовану і навчальні результати такого здобувача освіти потребують подальших досліджень.

Показано, що за допомогою розробленої програми можливо проводити кореляційний та регресійний аналіз, а також прогнозування, виявлення взаємозв'язків і викидів освітніх даних для вибірки великого розміру (1000 здобувачів освіти). Впровадження методів кореляції та регресії дає змогу автоматично обробляти результати моніторингу освітнього процесу, забезпечувати зворотні зв'язки у межах логічної та фізичної моделей системного аналізу якості підготовки педагогів професійного навчання.

У випадку складних залежностей між величинами для прогнозування їх значень є перспективним застосування штучних нейронних мереж. ШНМ можуть застосовуватися для встановлення залежностей не тільки між двома, але й багатьма величинами.

Список використаних джерел

1. Baker R., Yacef K. The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*. 2009. Vol. 1, № 1. P. 3–17. DOI: <https://doi.org/10.5281/zenodo.3554658>

2. Derevyanchuk O., Hu Z, Balovsyak S., Holub S., Kravchenko H., Sapsai I. Complex of Specialized Methods of Educational Data Mining for the Training of Vocational Education Teachers. *International Journal of Modern Education and Computer Science (IJMECS)*. 2025. Vol. 17, № 1. P. 28–46. DOI: <https://doi.org/10.5815/ijmeecs.2025.01.03>



3. Hu Z., Derevyanchuk O., Balovsyak S., Ushenko Y., Kravchenko H., Sapsai I. Fuzzy Clustering of Educational Data with Automated Selection of Processing Parameters in System Analysis of Quality Education. *International Journal of Modern Education and Computer Science (IJMECS)*. 2025. Vol. 17, № 4. P. 101–111. DOI: <https://doi.org/10.5815/ijmeecs.2025.04.07>
4. Згуровський М. З., Панкратова Н. Д. Основи системного аналізу. К. Видавнича група BHV, 2007. 544 с.
5. Грицюк П. М., Джоші О. І., Гладка О. М. Основи теорії систем і управління: навчальний посібник. Рівне : НУВГП, 2021. 272 с.
6. Barletta G., Trezza G., Chiavazzo E. Learning Effective Good Variables from Physical Data. *Machine Learning and Knowledge Extraction*. 2024. Vol. 6, № 3. P. 1597–1618. DOI: <https://doi.org/10.3390/make6030077>
7. Черняк О. І. Інтелектуальний аналіз даних: Підручник / О. І. Черняк, П. В. Захарченко. – Київський національний університет ім. Т. Шевченка. К. : Знання, 2014. 599 с.
8. Штовба С. Д. Інтелектуальні технології ідентифікації залежностей. Лабораторний практикум : електронний навчальний посібник / С. Д. Штовба, В. В. Мазуренко. – Вінниця : ВНТУ, 2014. 113 с.
9. Dao Thi Thanh Loan, Nguyen Duy Tho, Nguyen Huu Nghia, Vu Dinh Chien, and Tran Anh Tuan. Analyzing Students' Performance Using Fuzzy Logic and Hierarchical Linear Regression. *International Journal of Modern Education and Computer Science (IJMECS)*. 2024. Vol. 16, № 1. P. 1–10. DOI: <https://doi.org/10.5815/ijmeecs.2024.01.01>
10. Zhou Q., Quan W. et al. Predicting high-risk students using internet access logs. *Knowledge and Information Systems*. 2018. Vol. 55. P. 393–413. DOI: <https://doi.org/10.1007/s10115-017-1086-5>
11. Mukesh Kumar, Nidhi Walia, Sushil Bansal, Girish Kumar, Korhan Cengiz. Predicting College Students' Placements Based on Academic Performance Using



Machine Learning Approaches. *International Journal of Modern Education and Computer Science (IJMECS)*. 2023. Vol. 15, № 6. P. 1–13. DOI: <https://doi.org/10.5815/ijmeecs.2023.06.01>

12. Ahmad N., Hassan N., Jaafar H., and Enzai N. I. M. Students' Performance Prediction using Artificial Neural Network. *IOP Conf. Ser.: Mater. Sci. Eng.* 2021. Vol. 1176. P. 1–8. DOI: <https://doi.org/10.1088/1757-899X/1176/1/012020>

13. Rodríguez-Hernández C. F., Musso M., Kyndt E., Cascallar E. Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*. 2021. Vol. 2 P. 1–14, DOI: <https://doi.org/10.1016/j.caeai.2021.100018>

14. Sarsa S., Leinonen J., Hellas A. Empirical Evaluation of Deep Learning Models for Knowledge Tracing: Of Hyperparameters and Metrics on Performance and Replicability. *Journal of Educational Data Mining*. 2022. Vol. 14, № 2. DOI: <https://doi.org/10.5281/zenodo.7086179>

15. Дерев'янчук О. В. Оптимізація концептуальної моделі системного аналізу якості підготовки педагогів професійного навчання за допомогою інтелектуальних сценаріїв. *Педагогічна Академія: наукові записки*. 2025. № 18. DOI: <https://doi.org/10.5281/zenodo.15790424>

16. Дерев'янчук О. В. Фізична модель системного аналізу якості підготовки педагогів професійного навчання. *Педагогічна Академія: наукові записки*. 2025. № 20. DOI: <https://doi.org/10.5281/zenodo.15875126>

17. Гнеденко Б. В. Курс теорії ймовірностей. Київ : ВПЦ Київський університет, 2010. 464 с.

18. Карташов М. В. Імовірність, процеси, статистика. Київ : ВПЦ Київський університет, 2007. 504 с.

19. Balovsyak S., Derevyanchuk O., Kravchenko H., Ushenko Y., Hu. Z. Clustering Students According to their Academic Achievement Using Fuzzy Logic.



ПЕДАГОГІЧНА АКАДЕМІЯ:
НАУКОВІ ЗАПИСКИ

International Journal of Modern Education and Computer Science (IJMECS). 2023.

Vol. 15, № 6. P. 31–43. DOI: <https://doi.org/10.5815/ijmeecs.2023.06.03>