

 <https://doi.org/10.31651/2524-2660-2024-2-69-73>

 <https://orcid.org/0000-0002-2339-1994>


ГУРАЛЬ Інеса

кандидатка фізико-математичних наук, доцентка, доцентка кафедри фізико-математичних наук,
Івано-Франківський національний технічний університет нафти і газу
e-mail: inesa.gural@gmail.com

 <https://orcid.org/0000-0001-6100-1654>

ОСИПЧУК Михайло

доктор фізико-математичних наук, професор,
професор кафедри математичного і функціонального аналізу,
Прикарпатський національний університет ім. Василя Стефаника
e-mail: mykhailo.osypchuk@pnu.edu.ua

 <https://orcid.org/0000-0003-4451-4508>

СМОЛОВИК Ліана

кандидатка технічних наук, доцентка, доцентка кафедри фізико-математичних наук,
Івано-Франківський національний технічний університет нафти і газу
e-mail: lil02smo@gmail.com

УДК 378.015.3-057.87::51]:303.7(045)

СТАТИСТИЧНІ МЕТОДИ КЛАСИФІКАЦІЇ НАВЧАЛЬНОЇ ПОВЕДІНКИ СТУДЕНТІВ ПРИ ВИВЧЕННІ МАТЕМАТИКИ

Проведено кластерний аналіз мотиваційних переконань та навчальної діяльності майбутніх інженерів при вивченні дисциплін математичного циклу. Всі розрахунки та графічні побудови виконані з використанням програмного середовища статистичних розрахунків R та середовища розробки RStudio.

За підсумками кластерного аналізу отримано чотири значущо різні профілі студентів щодо мотиваційних переконань, самоефективності, стратегій саморегуляції, академічних досягнень та проблем при вивченні математичних дисциплін.

Наведена інтерпретація отриманих результатів. Розуміння типології студентів може дати орієнтири для вибудовування програм консультування та підтримки студентів, а також для планування навчального процесу та вдосконалення змісту і методів технічної освіти.

Запропоновано метод класифікації студентів в одержані групи.

Ключові слова: кластерний аналіз; дискримінантний аналіз; математика; мотивація; навчальна поведінка; профілювання студентів.

Метою дослідження було визначено виявлення кластерів студентів на основі їх мотивації та навчальної поведінки у вищому навчальному закладі при вивченні дисциплін математичного циклу.

Постановка проблеми в загальному вигляді. Математичний цикл дисциплін є важливою складовою освіти майбутнього інженера, оскільки його фахова підготовка значною мірою базується на математиці. Процес навчання в вищому навчальному закладі в зв'язку з реалізацією нових освітніх програм і посиленню ролі та постійної оптимізації самостійної роботи студентів визначається індивідуальними особливостями суб'єктної позиції студентів у навчанні та професійному розвитку. Численні до-

слідження показують, що вивчення математики часто супроводжується проблемами, пов'язаними з розумінням. Ключовим фактором для навчальної діяльності можуть бути: емоційне ставлення до математики (мені подобається/не подобається математика), уявлення про можливість/неможливість досягти успіху в навчанні, очікувана цінність для майбутньої кар'єри. Розуміння особливостей характеристик ставлення до математики студентів, які спонукають їх прикладати достатньо зусиль до вивчення дисциплін математичного циклу та залишатися наполегливими, навіть незважаючи на труднощі та перешкоди, може допомогти викладачам застосовувати ефективні стратегії викладання. Таким чином важливим є визначення груп студентів зі схожим рівнем сформованості навчальної діяльності. Таке профілювання здобувачів освіти дозволяє викладачам краще розуміти саморегульоване навчання студентів, щоб вони могли застосовувати ефективні стратегії викладання для підвищення їх мотивації.

Аналіз досліджень та публікацій. Використання методів кластеризації визначає закономірності для опису груп студентів, створення індивідуальних траєкторій навчання, розробки стратегій викладання та навчання, орієнтованих на однорідні групи студентів. Наприклад, кластеризація профілів студентів використовується при розробці адаптивних навчальних інструментів (Mojarad, Essa & Baker, 2018), для визначення уподобань в стилі навчання (Pasinaa et al., 2019). Дослідження (Ng, Liu & Wang, 2016) зосереджено на розумінні мотивації учнів середньої школи і саморегульованому навчанні з математики та природничих

наук за допомогою кластерного аналізу. В ряді робіт проводять кластеризацію здобувачів освіти в онлайн форматі (Kaur, Krishan, 2013), зокрема в масових відкритих онлайн-курсах (Tzeng et al., 2021). Такі дослідження дозволили виявити поведінкові стратегії студентів під час онлайн навчання.

Виклад основного матеріалу дослідження. Об'єктом дослідження було обрано 194 студента бакалаврату інженерних спеціальностей Івано-Франківського національного технічного університету нафти і газу (ІФНТУНГ). Для визначення характеристик ставлення студентів до вивчення дисциплін математичного циклу було складено анонімні анкети, в яких пропонувалось оцінити за п'ятибальною шкалою від 1 – повністю не згоден до 5 – повністю згоден наступні характеристики ставлення: 1) внутрішню цінність математики – інтерес до неї; 2) корисну цінність для майбутніх освітніх та кар'єрних прагнень; 3) самоефективність – впевненість в здатності оволодіти знаннями та навичками з математичних дисциплін в університеті, 4) тривожність під час контрольних заходів з математичних дисциплін, 5) рівень знань, 6) складність для вивчення математичних дисциплін, 7) саморегулювання - достатність зусиль для вивчення математичних дисциплін.

Мета дослідження була досягнута з використанням таких статистичних методів класифікації, як кластерний аналіз та дискримінантний аналіз.

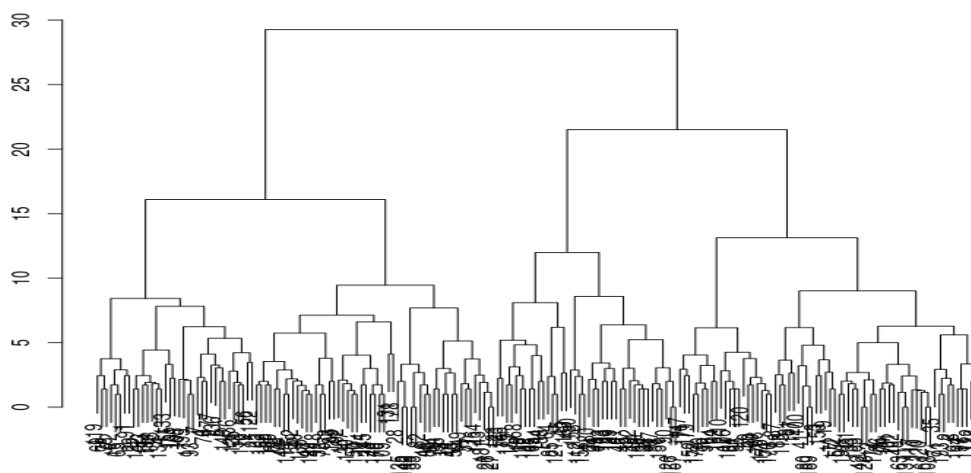


Рис. 1. Деревовидна діаграма.

Питання кількості кластерів, які слід розглядати, вирішувалось з використанням двовибіркового рангового критерія положення (Two Sample Rank Tests for Location based on Marginal Ranks) (див., наприклад, (Puri, Sen, 1971)). Досягнутий рівень значущості (p -value) обчислювався як максимум таких рівнів при перевірці гіпотез про

Визначення груп студентів, які максимально різні між собою та максимально гомогенні всередині, здійснено з допомогою метода деревовидної кластеризації. Кластерний аналіз є представником методів, так званої, класифікації без навчання. З його допомогою досліджувана вибірка ділиться на групи (кластери) за характеристиками її елементів.

Для практичного визначення належності студента до тієї чи іншої групи пропонується використати дискримінантний аналіз, який буде процедурою найкращого вибору групи для досліджуваного студента за його характеристиками. Цей метод відноситься до методів класифікації з навчанням. Першим кроком є вивчення особливостей груп, а другим побудова правил класифікації. Класифікація відбувається таким чином, щоб ймовірність помилки була найменшою можливою серед певного класу правил класифікації.

Всі розрахунки та графічні побудови в роботі виконані з використанням програмного середовища статистичних розрахунків R (R Core Team, 2024) та середовища розробки RStudio.

Кластеризація. Класифікація (поділ) студентів на групи виконана за допомогою кластерного аналізу. Застосована деревовидна кластеризація за методом Варда (Ward's method), використовуючи евклідову відстань (див., наприклад, Murtagh, Legendre, 2014). Деревовидна діаграма для такої кластеризації зображена на рис. 1.

рівність кожної характеристики в двох вибраних кластерах. Перебираючи кількість кластерів для розгляду від 2 і далі, порівнювали кожні два з них. Як тільки для двох якихось кластерів певного рівня гіпотеза про рівність характеристик підтверджувалась (чи не спростовувалась), вибирали для розгляду кількість кластерів

попереднього вищого рівня. Граничний рівень значущості було обрано рівним 0.05. Оскільки для чотирьох кластерів $p\text{-value}=0,00039$, а для п'яти $p\text{-value}=0,07926$, то слід обмежитися аналізом чотирьох кластерів.

Після побудови кластерної структури досліджуваної сукупності перевіряємо кожну характеристику на наявність відмін-

ностей в одержаних кластерах з допомогою критерію Краскела-Уоліса (Kruskal-Wallis Rank Sum Test) (див., наприклад, (Hollander, Wolfe, 1973)). Кожні два кластери порівнюємо з допомогою критерія Уїл-коксона. В таблиці 1 наведено $p\text{-value}$ парних порівнянь кластерів за кожною характеристикою.

Таблиця 1

Попарні порівняння кластерів за кожною характеристикою ($p\text{-value}$)

	Clusters 1&2	Clusters 1&3	Clusters 1&4	Clusters 2&3	Clusters 2&4	Clusters 3&4
Внутрішня цінність	0.0000	0.8744	0.0000	0e+00	0.0000	0.0000
Корисна цінність	0.2977	0.0158	0.0000	2e-04	0.0002	0.0000
Самоефективність	0.0000	0.9493	0.0000	0.0000	0.1541	0.0000
Тривожність	0.0000	0.0000	0.0332	0.0000	0.0000	0.0000
Рівень знань	0.0000	0.6360	0.0000	0.0000	0.4010	0.0000
Складність	0.0000	0.0011	0.0382	0.0000	0.0001	0.7361
Саморегулювання	0.0000	0.2284	0.0000	0.0000	0.0241	0.0000

Графічно відмінності в кластерах відображаються на коробчастих діаграмах (*box-and-whisker plot*) (рис.2). Ці діаграми відображають екстремальні (мінімальне та максимальне) значення, перший і третій квартилі, медіану.

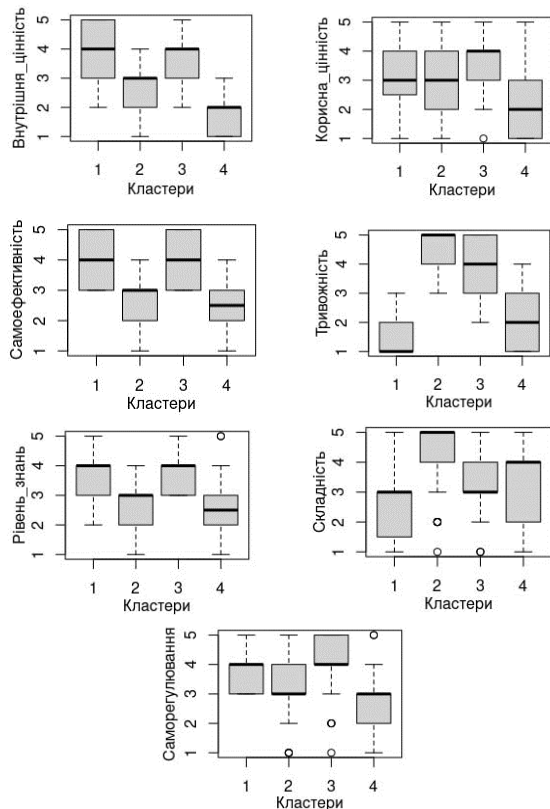


Рис. 2 Коробчасті діаграми для кожної з характеристик

Аналізуючи таблицю 1 і діаграми, значимо наступне. Студенти з 1-го кластеру

зацікавлені у вивченні математики: вони вважають її корисною для майбутніх освітніх та кар'єрних прагнень, в них достатньо високий рівень знань з математики і вони впевнені в здатності оволодіти знаннями та навичками з математичних дисциплін в університеті. При цьому в них найнижча тривожність під час контрольних заходів з математичних дисциплін і для них математичні дисципліни нескладні. Студентами другого кластеру математичні дисципліни сприймаються як нецікаві, складні. Вони до деякої міри розуміють корисну цінність цих дисциплін і прикладають якісь зусилля до навчання, найбільше хвилюються під час контрольних заходів. Але рівень знань, як і самоефективність у цих студентів невисокі. Студенти з третього кластеру прикладають найбільше зусиль для вивчення дисциплін математичного циклу, вони самоефективні і досягають хорошого рівня знань. Вони вважають математику скоріше цікавою і корисною для майбутніх освітніх та кар'єрних прагнень, хоча для них математичні дисципліни достатньо складні для опанування. Вони хвилюються під час контрольних заходів. Студенти з четвертого кластеру прикладають найменше зусиль для вивчення математичних дисциплін, які вважають нецікавими і некорисними. В них найнижчий рівень знань і самоефективність. Вони не переживають на контрольних заходах і не вважають математику складною. Однозначно студенти з цього кластеру не зацікавлені у навчанні.

Таким чином можна чітко виділити «позитивні» кластери (1 та 3) і «негативні» (2

та 4). Студенти з «позитивних» кластерів прикладають достатньо зусиль для вивчення дисциплін математичного циклу, розуміють їх цінність, мають хороший рівень знань, але студентам з першого кластеру легко опанувати ці дисципліни і вони не переживають на контрольних заходах на відміну від студентів з третього кластеру.

Група проблемних студентів також поділилася на два окремих кластери: студентам з другого кластеру найскладніше навчатися, вони найбільше хвилюються на контрольних заходах, але все ж докладають деяких зусиль, хоча й недостатніх (рівень знань невисокий), на відміну від студентів з четвертого кластеру, які є абсолютно немотивованими студентами з найнижчим саморегулюванням і рівнем знань.

Правило класифікації. Для класифікації нового об'єкта (студента) у побудовані та вивчені кластери можна запропонувати використати, так званий, лінійний дискримінантний аналіз. Цей метод аналізу даних проектує вектор характеристик на простір на одиницю меншої розмірності, ніж кількість груп (кластерів). Кожна координата проекції є значенням відповідної лінійної дискримінантної функції. Коефіцієнти таких функцій (LD1, LD2 та LD3) наведені в таблиці 2.

Таблиця 2

Коефіцієнти
лінійних дискримінантних функцій

Характеристики	LD1	LD2	LD3
Внутрішня цінність	-0.4467	-0.6615	0.9124
Корисна цінність	0.2088	-0.0733	-0.1501
Самоефективність	-0.4202	-0.2765	-0.3993
Тривожність	1.1880	-0.4876	-0.1886
Рівень знань	-0.1222	-0.3026	-0.7278
Складність	0.3133	0.0063	0.2811
Саморегулювання	-0.2959	-0.2356	0.0615

Використовуючи координати в просторі дискримінантних функцій центрів кластерів та середніх значень характеристик, можна визначити до якого кластера слід віднести досліджуваний об'єкт. Координати центрів кластерів наведені в таблиці 3, а середні значення характеристик – у таблиці 4.

Таблиця 3

Координати центрів кластерів

Кластери	LD1	LD2	LD3
1	-3.3048	-0.2834	0.2661
2	2.3878	0.0192	0.3499
3	0.3454	-1.4696	-0.2588
4	-0.0405	2.0010	-0.0294

Таблиця 4

Середні значення характеристик

Характеристики	Середні значення
Внутрішня цінність	2.9904
Корисна цінність	2.9572
Самоефективність	3.2133
Тривожність	3.0605
Рівень знань	3.0554
Складність	3.3316
Саморегулювання	3.3396

Приклад. Нехай характеристики деякого студента наступні

Внутрішня цінність	Корисна цінність	Самоефективність	Тривожність	Рівень знань	Складність	Саморегулювання
4	5	2	5	3	4	4

Обчислюємо відхилення характеристик від середніх значень.

Внутрішня цінність	Корисна цінність	Самоефективність	Тривожність	Рівень знань	Складність	Саморегулювання
1.0096	2.0428	-1.2133	1.9395	-0.0554	0.6684	0.6604

Проектуємо ці відхилення на простір дискримінантних функцій, помноживши вектор (рядок) відхилень на матрицю коефіцієнтів дискримінантних функцій.

LD1	LD2	LD3
2.8104	-1.5622	1.0021

Обчислюємо відстані (евклідові) від цієї проекції до центрів кластерів.

Кластер	Відстань
1	39.5735
2	3.1049
3	7.6748
4	21.8885

Найменше значення відстані вказує на номер кластера, до якого слід віднести досліджуваного студента. В нашому випадку, це – кластер номер 2.

Зауваження. Запропонований метод класифікації студентів на чотири групи дає не менше 90 відсотків правильних класи-

фікацій. В нашому дослідженні з використанням методу почергового тестування було одержано 91.24% правильних класифікацій.

Висновки та перспективи подальших досліджень. Використання кластерного аналізу в даному дослідженні виявилось плідним у визначенні підгруп студентів з різними моделями мотивації та навчальної поведінки при вивченні дисциплін математичного циклу. Це дозволяє викладачам, планувальникам навчальних програм, академічним менеджерам забезпечувати індивідуальне управління навчанням для різних профілів студентів та оптимізувати освітній процес з метою підвищення рівня знань. Також це може допомогти студентам зрозуміти власні сильні та слабкі сторони та шукати відповідні ресурси чи втручання.

Проведений аналіз базувався тільки на зборі даних виявлених власних переконань студентів, які дослідник не міг перевірити. Перспективи подальших досліджень пов'язуємо з різними інструментами збору даних, таких як спостереження та інтерв'ю, які дозволяють як кількісний, так і якісний аналіз. Це забезпечить насиченість і глибину висновків.

Список бібліографічних посилань

References

Hollander, Wolfe, 1973 – Hollander, M., Wolfe, D. (1973), *Nonparametric Statistical Methods*. New York: John Wiley & Sons. PP. 115–120.

- Kaur, Krishan, 2013 – Kaur, M., Krishan, K. (2013). Cluster Analysis of Behavior of E-learners. *International Journal of Soft Computing and Engineering (IJSCE)*, 3(2): 344-346.
- Mojarad, Essa & Baker, 2018 – Mojarad, S., Essa, A., Baker, R. (2018). Data-driven learner profiling based on clustering student behaviors: learning consistency, pace and effort. *In Intelligent Tutoring Systems: 14th International Conference, ITS 2018, Montreal, QC, Canada, June 11–15, Proceedings 14* (pp. 130-139).
- Murtagh, Legendre, 2014 – Murtagh, F., Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J Classif* 31: 274–295. <https://doi.org/10.1007/s00357-014-9161-z>.
- Ng, Liu & Wang, 2016 – Ng, B., Liu, W., Wang, J. (2016). Student Motivation and Learning in Mathematics and Science: A Cluster Analysis. *International Journal of Science and Mathematics Education*, 14: 1359–1376.
- Pasinaa et al., 2019 – Pasinaa, I., Bayrama, G., Labiba, W., Abdelhadia, A., Nurunnabi, M. (2019). Clustering students into groups according to their learning style. *Methods X* 6, pp. 2189–2197.
- Puri, Sen, 1971 – Puri, M., Sen, P. (1971), *Nonparametric Methods in Multivariate Analysis*, New York: Wiley.
- R Core Team, 2024 – R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Tzeng et al., 2021 – Tzeng, J.-W., Huang, N.-F., Lee, C.-N., Chen, Y.-H., Chuang, A.-C. (2021). Clustering analysis system based on students' motivation and Learning behavior in MOOC. *Journal of Liberal Arts and Social Sciences*, 17(2): 145–170.
- Venables, Ripley, 2002 – Venables, W., Ripley, B. (2002) *Modern Applied Statistics with S*. Fourth edition. Springer New York, NY. <https://doi.org/10.1007/978-0-387-21706-2>.

HURAL Inesa

Ph.D in Physical and Mathematical Sciences, Associate Professor,
Associate Professor of Higher Mathematics Department,
Ivano-Frankivsk National Technical University of Oil and Gas

OSYPCHUK Mykhailo

Doctor of Physical and Mathematical Sciences, Professor,
Professor at the Department of Mathematical and Functional Analysis,
Vasyl Stefanyk Precarpathian National University

SMOLOVYK Liana

Ph.D in Technical Sciences, Associate Professor, Associate Professor at Higher Mathematics Department,
Ivano-Frankivsk National Technical University of Oil and Gas

STATISTICAL METHODS OF CLASSIFICATION OF STUDENTS' EDUCATIONAL BEHAVIOR IN MATHEMATICS STUDY

Summary. A cluster analysis of motivational beliefs and educational activities of future engineers when studying the disciplines of the mathematical cycle was conducted. All calculations and graphical constructions were performed using the software environment for statistical calculations R and the development environment RStudio.

According to the results of the cluster analysis, four significantly different profiles of students were obtained regarding motivational beliefs, self-efficacy, self-regulation strategies, academic achievements and problems in studying mathematical disciplines. The interpretation of the obtained results is given.

Understanding the typology of students can provide guidelines for building counseling and support programs for students, as well as for planning the educational process and improving the content and methods of technical education. A method of classifying students into the resulting groups is proposed.

Keywords: cluster analysis, discriminant analysis, mathematics, motivation, learning behavior, student profiling.

Одержано редакцією 03.06.2024
Прийнято до публікації 15.06.2024