

**Завада О. П.**

*кандидат фіз.-мат. наук, доцент кафедри ІС у менеджменті*

*Львівський національний університет імені Івана Франка*

**Мартин О. М.**

*кандидат економічних наук, доцент кафедри права та менеджменту*

*Львівський державний університет безпеки життєдіяльності*

*м. Львів, Україна*

## **СТАТИСТИЧНІ ДОСЛІДЖЕННЯ УСПІШНОСТІ СТУДЕНТІВ ЗА ДАНИМИ АВТОМАТИЗОВАНИХ НАВЧАЛЬНИХ СИСТЕМ**

Згідно Національної стратегії розвитку освіти в Україні одним із пріоритетних напрямів в освіті є створення системи моніторингу якості [1, с. 5].

На даний час інформація для такого моніторингу зберігається на машинних носіях в різних автоматизованих навчальних системах, в таблицях звітності, на сайтах навчальних закладів. Для обґрунтованого прийняття управлінських рішень в навчальному процесі потрібно всесторонньо аналізувати цю інформацію. Опрацювання великих масивів даних проте є дуже громіздкою роботою. Тому дослідження як успішності студентів, так і навчального процесу в цілому реально можна виконувати лише із застосуванням як класичних, так і нових статистичних методів та програмних засобів аналізу даних.

Повноцінний статистичний аналіз успішності учнів (з використанням рангової кореляції) виконано в [2, с. 142]. Щоправда для цього була створена спеціальна програма [2, С. 139]. Багатовимірний статистичний аналіз успішності з використанням пакету SPSS виконувався в [3, с. 31].

Типові задачі дейтамайнінгу для аналізу даних в освіті та відповідні інструментарії описані в [4, с. 153]. Проте більшість із цих інструментаріїв не є доступними працівникам більшості навчальних закладів.

Для статистичного аналізу успішності студентів Львівського національного університету імені Івана Франка використовувалася інформація із систем Деканат та MOODLE.

На основі даних системи Деканат були вибрані оцінки 60 студентів по 14 предметах. Засобами Data Mining були розраховані частоти  $p(i,j)$  настання кожної з подій  $[i;j] = [\text{предмет}(i); \text{оцінка}(j)]$  ( $i=1, \dots, 60; j=1, \dots, 14$ ), а також частоти  $p(i_1, j_1, i_2, j_2)$  одночасного настання кожної пари подій  $[i_1; j_1; i_2; j_2] = [\text{предмет}(i_1); \text{оцінка}(j_1); \text{предмет}(i_2); \text{оцінка}(j_2)]$ . Було виявлено 32 такі пари, в яких частота  $p(i_1, j_1, i_2, j_2)$  одночасного їхнього настання є суттєво вищою, ніж добуток частот  $p(i_1, j_1)$  та  $p(i_2, j_2)$  настання подій  $[i_1; j_1]$  та  $[i_2; j_2]$  незалежно одна від одної. В практиці Data Mining розраховують величину

$$Lift[i_1; j_1; i_2; j_2] = \frac{p(i_1, j_1, i_2, j_2)}{p(i_1, j_1) \cdot p(i_2, j_2)} \quad (1)$$

При значеннях  $Lift[i_1; j_1; i_2; j_2] > 1$  правила  $\text{if } [i_1; j_1] \text{ then } [i_2; j_2]$  та  $\text{if } [i_2; j_2] \text{ then } [i_1; j_1]$  є корисними (представляють інтерес) для подальших досліджень. Виявилось, що найбільші значення величини Lift мали такі пари предметів: [вища математика; оцінка; мікроекономіка; ця ж оцінка], а також [теорія ймовірності; оцінка; макроекономіка; ця ж оцінка]. Значення Lift для пар предметів [вища математика та теорія ймовірності] і [мікроекономіка; макроекономіка] серед високих не були. Таким чином застосування методів Data Mining виявило раніше не передбачувані зв'язки між предметами. Отримані результати можна змістовно пояснити таким чином: задачі мікроекономіки справді є дуже подібними до математичних задач на дослідження функцій. В той же час як предмет теорія ймовірностей, так і предмет макроекономіка вимагають від студента володіти мисленням на рівні змістовного моделювання.

Застосування як класичних статистичних методів, так і методів Data Mining на основі даних системи Деканат стало можливим тому, що ці дані зберігаються в EXCEL-таблицях і звідси можуть бути експортовані в спеціалізовані пакети. Щоправда дані довелося консолідувати із багатьох таблиць в одну.

На основі даних системи MOODLE виконувався аналіз результатів тестування з предмету “інформаційно-комунікаційні технології” для студентів першого курсу. Всього проходило тестування біля 600 студентів. Максимальний

бал становив 50. Система MOODLE підрахувала частоту  $p(i)$  для кожного набраного балу  $i$  ( $i = 10, \dots, 50$ ). Отриманий дискретний ряд

$$p_{10}, p_{11}, \dots, p_{50} \quad (2)$$

був досліджений на коефіцієнти варіації, асиметрії і ексцесу.

Середній бал тестування становив  $\mu = 36,0$ . Враховуючи той факт, що отримання 25 балів вважається незадовільним, а серединою інтервалу  $[25 - 50]$  є значення 37,5, то можна вважати, що з цієї позиції набір тестових завдань був не занадто складним і не занадто легким.

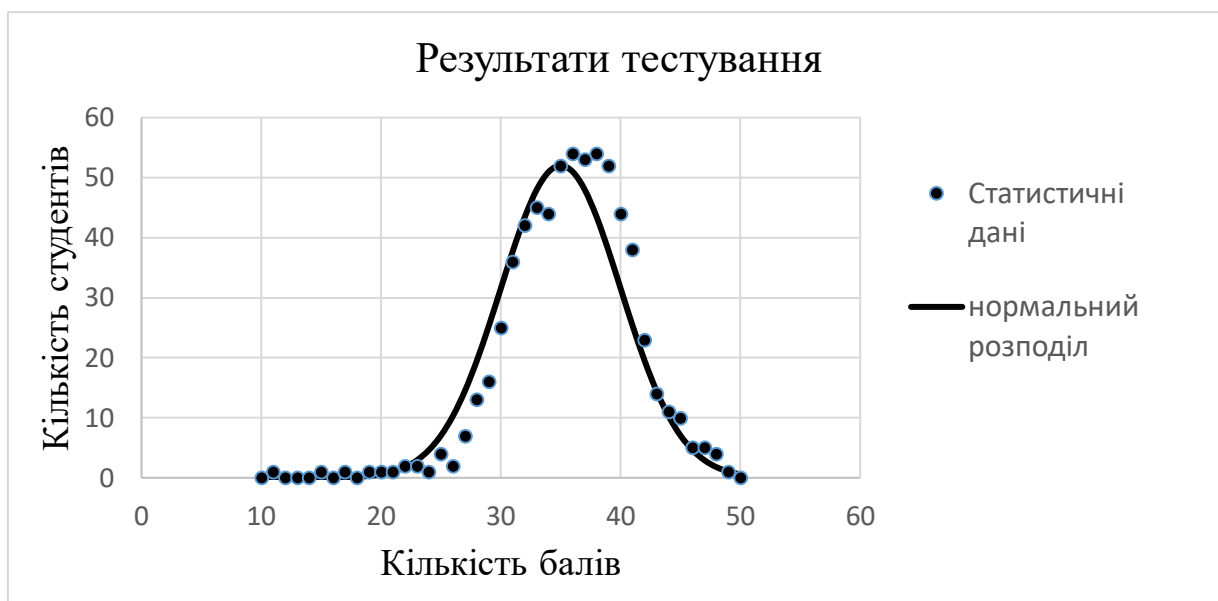
Коефіцієнт варіації  $V = 13,5\%$ , який був обчислений як процентне відношення середнього квадратичного відхилення до середнього арифметичного, можна вважати помірним. Значення величин  $\mu$  та  $V$  свідчать про досить вдалу побудову тестових завдань.

Коефіцієнт асиметрії  $A = -0,34$  виявився від'ємним, тобто в розподілі (2) спостерігалася лівобічна асиметрія [5, с. 43], якої би не мало бути. Бачиться таке пояснення: в тестах була занадто велика кількість запитань, на які студент міг правильно відповісти навіть не вивчаючи матеріалу курсу.

Коефіцієнт ексцесу ряду (2) становив  $E = 0,81$ . Це означає, що вершина розподілу тестових балів, отриманих студентами, знаходиться дещо вище, ніж тоді, коли цей розподіл був би нормальним. Цей факт можна пояснити тим, що викладалися нові теми, які ще не були належно методично забезпеченими. Тому було отримано занадто мало найвищих балів і занадто багато балів середніх.

Відомо, що нормальний розподіл виникає тоді, коли певна випадкова величина є сумою великої кількості незалежних випадкових величин, кожна з яких відіграє незначну роль в утворенні всієї суми. Асиметрія та ексцес є мірами відхилення від нормального розподілу. У розглянутому прикладі ці відхилення не є великими. Крім того точковий графік (рис. 1) статистичних даних не дуже відхиляється від графіку нормального розподілу з отриманими значеннями середнього та дисперсії. Тому можна вважати [6, с. 140], що результати тестування мають розподіл, близький до нормального, а сама система тестування була розроблена достатньо професійно.

Варто проте зазначити, що результати тестування із системи MOODLE в систему EXCEL доводилося переносити вручну.



**Рис. 1. Графік розподілу балів при тестуванні студентів**

*Побудовано на основі результатів тестування студентів економічного факультету*

Можна зробити низку висновків. В навчальному процесі для прийняття науково обґрунтованих рішень потрібно виконувати глибинний аналіз даних успішності. Ці дані знаходяться у файлах навчальних систем в достатній кількості. Проте практично всі навчальні системи лише автоматизують облік інформації. Вони дотепер є OLTP-системами. У них потрібно консолідувати дані з різних джерел та виконувати перетворення форматів даних. Утворити складний запит чи виконати серйозне аналітичне дослідження є можливим лише в тих випадках, коли викладач є програмістом або дуже кваліфікованим користувачем. На даний час з'являється деяке програмне забезпечення для аналізу успішності [7, с. 104] з використанням баз даних. Проте широке використання OLAP-систем в освіті буде можливим лише тоді, коли викладачі та керівництво навчальними закладами будуть знати можливості як класичних, так і сучасних статистичних методів. На жаль на даний час доводиться погоджуватися з [4, с. 163], де констатується факт, що "більшість українських дослідників нічого не знають ні про задачі, які можна ефективно розв'язувати за допомогою Data Mining, ні про наявний інструментарій". Хоча позитивним є той факт, що навчальні заклади, зокрема Львівський національний університет імені Івана Франка, розпочинають

проводити курси підвищення викладацької майстерності, де велика увага приділяється інформаційним технологіям та методам аналізу даних.

### **Список використаних джерел:**

1. Бодненко Д. М., Жильцов О. Б., Лещинський О. Л., Мазур Н. П. Моніторинг навчальної діяльності: навчальний посібник / Д. М. Бодненко, О. Б. Жильцов, О. Л. Лещинський, Н. П. Мазур — К: Київський університет імені Бориса Грінченка, 2014. — 276 с.
2. Селезньова Н. П. Кореляційний аналіз навчального процесу на прикладі підсумкових оцінок учнів / Н. П. Селезньова, Н. В Селезньова, С. В. Селезньов // Вісник НТУУ «КПІ». Філософія. Психологія. Педагогіка: збірник наукових праць. – 2012. – № 1(34). – С. 139–145.
3. Дериглазов Л. В., Кухаренко В. М., Перхун Л. П., Товмаченко Н. М. Статистичні методи аналізу результатів тестового контролю знань / Л. В. Дериглазов та ін. // MoodleMoot Ukraine 2018. Теорія і практика використання системи управління навчанням Moodle: тези доп. Шостої міжнар. наук.-практ. конф. Київ: КНУБА, 2018. - С. 28-32.
4. Ковальчук Ю. О. Пошук, отримання й аналіз даних в освіті: сучасний стан і перспективи розвитку /Ю. О. Ковальчук // Інформаційні технології і засоби навчання, 2015. Том 50. - №6. - С. 152 -163.
5. Кухаренко В. М., Перхун Л. П., Товмаченко Н. М. Методика комплексного оцінювання якості тестів. Частина 1 / В. М. Кухаренко, Л. П. Перхун, Н. М. Товмаченко // Статистика України, 2018. - № 3. - С. 40-48.
6. Кухар Л. О., Сергієнко В. П. Конструювання тестів. Курс лекцій: навч. посіб. / Л. О. Кухар, В. П. Сергієнко. – Луцьк, 2010. – 182 с.
7. Устенко С., Прадєдов В. Розробка програмного забезпечення для автоматизації обліку та аналізу поточної успішності студентів / С. Устенко, В. Прадєдов // Геометричне моделювання та інформаційні технології. - 2016. - № 1. - С. 103-109.