

НЕПАРАМЕТРИЧНІ МЕТОДИ МАТЕМАТИЧНОЇ СТАТИСТИКИ ДЛЯ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ПЕДАГОГІЧНИХ НОВОВВЕДЕНЬ

У статті розглянуто особливості непараметричних методів та їх використання в педагогічних дослідженнях; охарактеризовано конкретні види непараметричних критеріїв та принципи їх вибору для перевірки гіпотез.

Ключові слова: експеримент, гіпотеза, непараметричний критерій, вибірка, розподіл.

Використання математичних методів у педагогічних дослідженнях не завжди супроводжується чіткими уявленнями про можливості й межі їх застосування. Досить часто дослідник використовує той чи інший метод аналізу даних лише внаслідок його доступності. Основною метою використання математичних методів у педагогічних дослідженнях є створення формального математичного апарату для адекватного опису й моделювання систем, що мають певні властивості, та вираження закономірностей і залежностей у компактній формі.

Методи непараметричної статистики – порівняно нова математична дисципліна, яка нині активно розвивається. Універсальність, простота і точність непараметричних гіпотез привертають усе більше уваги дослідників. Застосування цих методів у педагогічних дослідженнях значно підвищить їхню ефективність та якість.

Основи непараметричних методів математичної статистики було розроблено спеціально для цілей педагогічних і психологічних досліджень Ч. Гальтоном, П. Зигелем, Т. Келлі, Л. Терстоном, Е. Торндайком та ін. Як підкреслює Є.В. Сидоренко, чим простіші методи математичного опрацювання, тим більш надійними й усвідомленими будуть результати.

Мета статті – обґрунтування переваг для обробки кількісних результатів педагогічного експерименту непараметричних критеріїв порівняно з параметричними.

Педагогічний експеримент дає змогу перевірити ефективність нововведень у сфері освіти, порівняти вагомість різних факторів у структурі педагогічного процесу й обрати їх оптимальне поєднання. Це своєрідний комплекс методів дослідження, що забезпечує науково-об'єктивну й обґрунтовану перевірку істинності гіпотези, висунутої на початку дослідження.

Від дослідника педагогічний експеримент вимагає високої методологічної культури, ретельної розробки програми й надійного статистичного критерію. На думку В.С. Аванесова, специфіка педагогічних експериментів полягає в тому, що в них рідко виконується умова випадковості відбору. Студенти працюють в усталених групах, і в процесі навчання важко відібрати у випадковому порядку студентів із різних груп та різних навчаль-

них закладів, а потім сформувані з них експериментальну й контрольну вибірку. Ця обставина змушує вважати такі експерименти так званими квазіекспериментами, тобто неточними [1].

Як вихід з описаної ситуації можна використовувати “план з нееквівалентною контрольною групою” за класифікацією Д. Кемпбелла, оскільки в педагогічних дослідженнях значного поширення набув експеримент, у якому над експериментальною й контрольною групами проводиться тестування до та після педагогічного впливу. При цьому обидві групи нееквівалентні одна одній до здійснення педагогічного впливу. Такими групами можуть бути дві паралельні групи на потоці або два потоки на курсі [6].

Головне завдання науки – це формулювання й перевірка гіпотез. Із цієї точки зору експеримент є засобом контролю, завдяки якому відхиляються недостовірні припущення.

Дослідники виділяють наукові та статистичні гіпотези. Наукова гіпотеза формулюється як гіпотетичний розв’язок проблеми. Статистична гіпотеза – твердження стосовно невідомого параметра, сформульоване мовою математичної статистики. Кожна наукова гіпотеза потребує перекладу на мову статистики. Під час організації експерименту кількість гіпотез обмежується до двох: основної та альтернативної. Процедура статистичної інтерпретації даних зводиться до оцінювання еквівалентностей і розбіжностей. При перевірці статистичних гіпотез використовують лише два поняття: H_0 (гіпотеза про збіжність, у якій дослідник припускає, що нова методика, новий метод, прийом чи технологія не має якихось переваг, і тому від самого початку дослідник готовий зайняти наукову позицію “різниця між новою і старою методикою дорівнює нулю”) та H_1 (гіпотеза про розбіжність, у якій робиться припущення про переваги нової методики). Підтвердження альтернативної гіпотези свідчить про правильність статистичного твердження H_1 , а підтвердження основної – про відсутність розбіжностей. Саме альтернативні гіпотези відіграють головну роль в алгоритмі оцінювання ефективності педагогічних нововведень при застосуванні непараметричних методів математичної статистики [4; 8].

Д. Гласс і Д. Стенлі наголошують, що при перевірці довільної статистичної гіпотези висновок дослідника оцінюється не в сто відсотків, оскільки завжди допускається ризик прийняття неправильного рішення. Заздалегідь задається достатньо мала ймовірність α (наприклад, рівна 0,1; 0,05; 0,01), яку називають рівнем значущості. Відповідно, якщо подія з імовірністю, що не перевищує α , все-таки відбулася, то настання такої події не може бути викликано лише випадковими причинами. Цю подію варто розглядати як “невипадкову”. Вона повинна привернути увагу дослідника і стати для нього “значущою”. При цьому чим менша ймовірність здійснення події, тим більша її “невипадковість”, і тим стає важливішим встановити її причини [3].

Якщо критерій ґрунтується на певному типі розподілу генеральної сукупності або використовує параметри цієї сукупності, то такий критерій називають параметричним. Вивчають сім’ї нормальних розподілів, логариф-

мічно нормальних, експоненційних, гамма-розподілів тощо. Усі вони залежать від одного, двох або трьох параметрів. Тому для повного опису розподілу досить знати або оцінити одне, два або три числа. Параметрична теорія математичної статистики передбачає, що розподіл результатів спостережень належить тим чи іншим параметричним сім'ям. Найбільш популярний нормальний розподіл. Критерії, що ґрунтуються на t -розподілі Стюдента і F -розподілі Фішера – Снедекора, впливають із припущення про нормальний закон розподілу випадкових величин генеральної сукупності. Іноді пропонують перед застосуванням, наприклад, критерію Стюдента однорідності двох вибірок перевіряти нормальність. Хоча для цього є багато критеріїв, але перевірка нормальності – більш складна і трудомістка статистична процедура, ніж перевірка однорідності (за допомогою непараметричних критеріїв). Для достатньо надійного встановлення нормальності потрібна дуже велика кількість спостережень. Так, щоб гарантувати, що функція розподілу результатів спостережень відрізняється від деякої нормальної не більше ніж на 0,01, потрібно близько 2500 спостережень. У більшості педагогічних досліджень кількість спостережень істотно менша. При цьому, розподіл реальних даних майже ніколи не входить до конкретної параметричної сім'ї. Реальні розподіли завжди відрізняються від тих, що включені до параметричної сім'ї. Звідси впливає необхідність вивчення стійких і непараметричних, у тому числі вільних від розподілу процедур, їх широкого впровадження в практику статистичної обробки даних [2; 7].

Якщо критерій не спирається на припущення про тип розподілу генеральної сукупності й не використовує параметри цієї сукупності, то він називається непараметричним. При обробці реальних даних слід застосовувати стійкі методи, зокрема непараметричні.

Непараметричні методи математичної статистики застосовують для перевірки гіпотез таких типів:

1. Дві (або більше) властивості досліджуваного педагогічного явища стохастично залежні, і ця залежність описується певним законом (наприклад, лінійним); деякий фактор (фактори) впливає на досліджувану властивість педагогічного явища, і ця стохастична залежність описується певним законом. У такому разі можливі якісні вимірювання (коефіцієнти Кандела, Спірмена тощо). Але більш глибокий аналіз стохастичних зв'язків, а саме встановлення щільності й типу зв'язку, здійснюється за допомогою методів кореляційного, регресійного та дисперсійного аналізів. Їх можна провести лише на основі кількісних вимірювань.

2. Стан однієї й тієї самої властивості має однаковий чи різний розподіл у кожній із двох сукупностей студентів, які відрізняються змістом, методом, організацією навчального процесу або соціальним середовищем. Перевірку гіпотези такого типу проводять за допомогою методів, що дістали назву критеріїв значущості. Частина з них використовується лише при кількісному вимірюванні досліджуваної властивості, інші ж можна використовувати і при якісному вимірюванні.

Можливості практичного застосування параметричних і непараметричних критеріїв математичної статистики суттєво залежать від вимірювань, які може провести дослідник. Параметричні методи для свого використання потребують вимірювання за інтервальною шкалою або шкалою відношень. Непараметричні ж методи можна застосовувати й у випадках, коли результати виміряні за шкалою найменувань або порядковою шкалою. Шкала найменувань дає змогу відрізнити один об'єкт від іншого; порядкова шкала класифікує за принципом “більше – менше”; інтервальна шкала відрізняє об'єкти за принципом “більше або менше на певну кількість одиниць”; шкала відношень – “у скільки разів більше або менше”. За допомогою перших двох шкал проводять якісні вимірювання, за третьою і четвертою – кількісні.

За словами М.І. Грабар, перевага непараметричних методів полягає у можливості їх використання для вибірок невеликого обсягу, а також у відносній простоті обчислення, пов'язаній із практичним застосуванням цих методів [5].

Для успішного практичного застосування непараметричних методів необхідно дотримуватися таких рекомендацій:

а) мету дослідження потрібно обрати таким чином, щоб забезпечити порівняння двох серій вимірювання (вибірок). Ці вибірки можуть бути залежними одна від одної або незалежними. Наприклад, вимірювання рівня знань у групі студентів до застосування певного педагогічного впливу і після його застосування будуть залежними, а такі самі вимірювання у двох різних студентських групах – незалежними;

б) вибірка результатів із загальної кількості вимірювань повинна бути випадковою і не мати суб'єктивного чи закономірного характеру;

в) нульова та альтернативна гіпотези повинні бути чітко сформульовані, мати однозначний характер. Формулювання гіпотез повинно відповідати обраному для застосування критерію;

г) якщо оцінюють наявність чи відсутність певної властивості, то доцільно наявність властивості позначити через “1”, а відсутність – через “0”;

д) при оцінюванні результатів перевірки гіпотез необхідно орієнтуватися на частоти виникнення подій, отриманих у процесі вимірювання. У випадках, коли частоти різко відрізняються, застосування методів перевірки гіпотез недоцільне, оскільки результат очевидний;

е) результат після перевірки гіпотези повинен бути інтерпретований певним чином.

М.І. Грабар описав сім найбільш важливих непараметричних критеріїв, які варто застосовувати в педагогічних дослідженнях:

1. Критерій Макнамара дає змогу порівнювати педагогічні об'єкти за зміною частоти певної властивості при двох залежних вимірюваннях до застосування педагогічного впливу і після нього. Для цього критерію мінімальний обсяг вибірки – 10–15 осіб.

Формують дві гіпотези. Нульова гіпотеза H_0 припускає, що досліджувана властивість після застосування педагогічного впливу не змінилася.

Альтернативна гіпотеза H_1 припускає, що досліджувана властивість змінилася (збільшилася або зменшилася).

Значення властивості при першому вимірюванні позначається через $X = \{x_1, x_2, \dots, x_i, \dots, x_N\}$, при повторному вимірюванні – $Y = \{y_1, y_2, \dots, y_i, \dots, y_N\}$, де $i = 1, N$ – індекс вимірюваного об'єкта; $x_i, y_i = \{0, 1\}$, де “0” – відсутність властивості, що вимірюється, а “1” – наявність цієї властивості.

На основі векторів X і Y формулюють статистику за такими формулами: $S1_i = X_i \wedge \overline{Y_i}$, $S2_i = \overline{X_i} \wedge Y_i$; $Q_1 = \sum_{i=1}^N S1_i$, $Q_2 = \sum_{i=1}^N S2_i$.

Якщо $Q_1 = Q_2$, то критерій Макнамара не використовують. Крім того, має виконуватися умова $n = Q_1 + Q_2 \geq 5$.

Якщо $Q_1 > Q_2$, то досліджувана властивість зростає, якщо ж $Q_1 < Q_2$, то досліджувана властивість зменшується.

Для обчислення значення статистики застосовують формулу:

$$T_{\text{емп}} = \begin{cases} \min(Q_1, Q_2), & \text{якщо } Q_1 + Q_2 \leq 20 \\ \frac{(Q_1 - Q_2)^2}{Q_1 + Q_2}, & \text{якщо } Q_1 + Q_2 > 20 \end{cases}$$

2. Критерій знаків дає змогу порівняти педагогічні об'єкти на основі аналізу двох залежних вимірювань. Для цього критерію мінімальний обсяг вибірки – 15–20 осіб. Величину рівня значущості α варто обрати 0,05; 0,02; 0,01 або 0,025; 0,01; 0,005 залежно від варіанта обраного критерію (односторонній чи двосторонній).

За початкові дані беруть вектор $X = \{x_1, x_2, \dots, x_i, \dots, x_N\}$, отриманий шляхом вимірювання досліджуваної властивості до застосування педагогічного впливу, і вектор $Y = \{y_1, y_2, \dots, y_i, \dots, y_N\}$ – після педагогічного впливу. Пари X_i, Y_i повинні бути незалежними між собою і набувати значення цілих чисел. Це можуть бути бали, присвоєні в процесі вимірювання i -му об'єкту. Прикладом може бути традиційна 5-бальна система оцінювання студентів.

Окремі пари (X_i, Y_i) , де i – номер вимірювання в першому і другому випадках, аналізуються і розбиваються на три класи. Якщо $X_i < Y_i$, то парі присвоюється число “1”, якщо $X_i > Y_i$ – число “-1”, якщо ж $X_i = Y_i$, то присвоюється значення “0”.

Формулюють нульову гіпотезу $H_0: (X_i < Y_i) = P (X_i > Y_i)$ для всіх $i = \overline{1, N}$. Зміст цієї гіпотези в тому, що рівність $P (X_i < Y_i)$ і $P (X_i > Y_i)$ свідчить про рівність законів розподілу випадкових величин, отриманих при першому і другому вимірюваннях. Як альтернативна гіпотеза H_1 виступає нерівність цих законів. $H_1: P (X_i < Y_i) \neq P (X_i > Y_i)$ для всіх $i = \overline{1, N}$.

Далі знаходять значення Z за правилом: $Z_i = “0”$, якщо $X_i = Y_i$; $Z_i = “+1”$, якщо $X_i < Y_i$; $Z_i = “-1”$, якщо $X_i > Y_i$. Емпіричне значення статистики $T_{\text{емп}}$ дорівнює кількості значень $Z_i = “+1”$: $T_{\text{емп}} = \sum_{i=1}^N (Z_i = +1)$.

Визначення $T_{\text{емп}}$ проводять по-різному для випадку, коли $n \leq 100$ і $n > 100$. Коли $n \leq 100$, потрібно скористатися спеціально розробленою таблицею, вхідними параметрами для якої будуть n і α . У результаті отримаємо величину t_α і матимемо змогу обчислити значення $T_{\text{крит1}} = t_\alpha$ і $T_{\text{крит2}} = n - t_\alpha$.

Двосторонній критерій безпосередньо не дає відповіді про покращення чи погіршення досліджуваної властивості.

Лівосторонній критерій передбачає тенденцію перевищення в середньому X_i над Y_i . Гіпотези $H_0: P(X_i < Y_i) \geq P(X_i > Y_i)$ та $H_1: P(X_i < Y_i) < P(X_i > Y_i)$. Нульову гіпотезу в цьому випадку відхиляють, якщо $T_{\text{емп}} < T_{\text{крит1}}$.

Правосторонній критерій передбачає тенденцію перевищення в середньому Y_i над X_i . Гіпотези $H_0: P(X_i < Y_i) \leq P(X_i > Y_i)$ та $H_1: P(X_i < Y_i) > P(X_i > Y_i)$. Нульову гіпотезу відхиляють, якщо $T_{\text{емп}} \geq T_{\text{крит2}}$.

Як зазначає Є.В. Сидоренко, критерій знаків дає змогу встановити, чи змінюються показники в бік зростання або, навпаки, зниження, тобто критерій знаків відповідає принципу “більше-менше”.

3. Критерій Вілкоксона (шкала вимірювання – інтервальна, пропорційна). Якщо зіставляють два вимірювання експериментальної й контрольної вибірок, причому опрацювання обох вибірок незалежне, то доцільно обрати критерій Вілкоксона. Використовуючи його, можна говорити не лише про підвищення або зниження показника, а й про те, на скільки відбулося таке підвищення або зниження.

4. Медіанний критерій дає можливість порівнювати педагогічні об’єкти на основі аналізу двох серій незалежних вимірювань (вибірок). Обсяг вибірки повинен перевищувати 20.

Перевірка полягає у виділенні розбіжностей центральної тенденції вибірок. Як показник виступає медіана досліджуваної властивості в кожній із вибірок. Медіана – величина ознаки, що знаходиться посередині ранжованого ряду вибірки, тобто це величина, яка знаходиться в середині ряду величин, розташованих у зростаючому або спадному порядку.

Обчислення емпіричного значення критерію здійснюється за формулою:

$$T_{\text{емп}} = \frac{(N_1+N_2) \left(|Q_{1x} \cdot Q_{2y} - Q_{2x} \cdot Q_{1y}| - (N_1+N_2)/2 \right)^2}{(Q_{1x} + Q_{1y})(Q_{2x} + Q_{2y})(Q_{1x} + Q_{2x})(Q_{1y} + Q_{2y})}$$

Критичне значення критерію $T_{\text{крит}}$ знаходиться за таблицею розподілу χ^2 (табл. 1).

Таблиця 1

α або $1-\alpha$	0,1 або 0,9	0,05 або 0,95	0,025 або 0,975	0,01 або 0,99
$T_{\text{крит}}$	2,706	3,841	5,024	6,635

5. Критерій Вілкоксона-Манна-Уїтні (ВМУ) (двосторонній і односторонній, шкала вимірювання – порядкова, інтервальна, пропорційна).

6. Критерій Хі-квадрат (двосторонній і односторонній, шкала вимірювання – шкала найменувань з декількома категоріями: “погано, задовільно, добре, відмінно” тощо).

7. Критерій Колмогорова – Смірнова (двосторонній і односторонній, шкала вимірювання – порядкова, інтервальна, пропорційна) [5].

Критерій Хі-квадрат дає змогу використовувати до ста категорій (100-бальну шкалу, що в наш час є актуальним, оскільки навчання у вищих навчальних закладах проводиться за Болонською системою). Цей критерій варто використовувати при порівнянні результатів двох залежних вибірок (повторне вимірювання однієї й тієї самої властивості або різних властивостей у членів однієї й тієї самої групи студентів) [9].

Для успішного застосування непараметричних методів математичної статистики при оцінюванні ефективності педагогічних нововведень досліднику необхідне розуміння не лише суті обраних методів, а й педагогічного експерименту, квазіексперименту, термінів “рандомізація”, “репрезентативність”, “валідність”, чотирьох типів шкал вимірювання, нульової й альтернативної гіпотез, рівнів значущості та достовірності, а також правильне практичне застосування цих теоретичних положень. Побудувавши гіпотези про досліджуване педагогічне явище, зробивши певні припущення, дослідник може скористатися широким арсеналом прийомів і методів математичної статистики. При цьому, якщо в педагогічному дослідженні користуватися ними коректно, то не повинно виникнути сумнівів стосовно справедливості отриманих висновків, навіть якщо немає глибокого розуміння всіх формальних математичних перетворень.

Висновки. Популярність непараметричних методів пояснюється стійкістю висновків, простотою математичних засобів. Непараметричні критерії значно менш трудомісткі, а при розподілах, далеких від нормального, більш ефективні й точні, ніж параметричні.

Список використаної літератури

1. Аванесов В.С. Применение статистических методов в педагогических исследованиях / В.С. Аванесов // Введение в научное исследование по педагогике / под ред. В.И. Журавлева. – М., 1988. – С. 16–21.
2. Боровков А.А. Математична статистика / А.А. Боровков. – М. : Наука, 1984. – 472 с.
3. Гласс Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стэнли. – М. : Прогресс, 1976. – 495 с.
4. Гончаренко С.У. Педагогічні дослідження / С.У. Гончаренко. – К., 2008. – 278 с.
5. Грабарь М.И. Применение математической статистики в педагогических исследованиях: Непараметрические методы / М.И. Грабарь, К.А. Краснянская. – М. : Педагогика, 1977. – 136 с.
6. Кэмпбелл Д. Эксперименты и квазиэксперименты. Модели экспериментов в социальной психологии и прикладных исследованиях / Д. Кэмпбелл. – М. : Прогресс, 1980. – С. 34–48.
7. Ительсон Л.Б. Математические и кибернетические методы в педагогике / Л.Б. Ительсон. – М. : Просвещение, 1964. – 268 с.

8. Крамер Г. Математические методы статистики / Г. Крамер. – М. : Мир, 1975. – 648 с.
9. Сидоренко Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб. : Речь, 2000. – 350 с.

Стаття надійшла до редакції 02.09.2013.

Криворот Т.Г. Непараметрические методы математической статистики для оценки эффективности педагогических новшеств

В статье рассмотрены особенности непараметрических методов и их использование в педагогических исследованиях; дана характеристика непараметрических критериев, описаны принципы их выбора для проверки гипотез.

Ключевые слова: эксперимент, гипотеза, непараметрический критерий, выборка, распределение.

Krivorot T. Non-parametric statistical methods to assess the effectiveness of pedagogical innovations

The article discusses the features of nonparametric methods and their use in educational research, the characteristic of non-parametric tests, describes the principles of their choice to test hypotheses.

The main purpose of the use of mathematical methods in educational research is the creation of formal mathematical tools for adequate description and modeling systems possess certain properties, and expression patterns and dependencies in compact form, etc.

The use of mathematical methods in educational research, unfortunately, not always accompanied by a clear idea of the possibilities and limits of their application. Often researcher uses a particular method of data analysis, not because it is the most feasible, but only because of its affordability. As a result, for the majority of applications Studies using simplified one-dimensional analysis, which considerably detracts information reduces the diagnostic and prognostic value obtained conclusions.

Use of these types of analysis as a factor, multivariate, variance, regression is very complex. Meanwhile, there are mathematical methods that are still quite rarely used in educational research, although their use would spend quite correct mathematical processing of data without forcing the researcher assimilate complex sections of higher mathematics, and would greatly simplify the calculations and calculation. We refer to methods of nonparametric statistics – a relatively new mathematical discipline that is actively developing.

Popularity nonparametric methods due to resistance conclusions simplicity mathematical tools. Non-parametric criteria are much less labor intensive, and the distribution, far from normal, more efficient and accurate than parametric.

The successful application of nonparametric methods of mathematical statistics in assessing the effectiveness of pedagogical innovations researcher requires an understanding not only the substance of selected methods, but also pedagogical experiment, the terms “randomization”, “representation”, “validity” of four types of scales of measurement, zero and alternative hypotheses, levels of significance and reliability, and correct practical application of these theoretical positions. Constructing hypotheses about the studied phenomenon teacher by making certain assumptions, the researcher can use a wide arsenal of techniques and methods of mathematical statistics. Thus, if a pedagogical study to use them correctly, you should have no doubts about the fairness of these findings, even if there is no deep understanding of all formal mathematical transformations.

Key words: experiment, hypothesis, nonparametric test, sample, distribution.