

УДК 37.012:004.422.636.7

АВТОМАТИЗАЦІЯ ОБРОБКИ ДАНИХ СТАТИСТИЧНОГО АНАЛІЗУ В НАУКОВИХ ДОСЛІДЖЕННЯХ

Крамаренко Т. А.

AUTOMATION OF TREATMENT OF DATA OF STATISTICAL ANALYSIS IN RESEARCHES

Kramarenko T. A.

У статті розглянуто використання табличного процесору Excel для автоматизації розрахунку емпіричного значення критерію однорідності χ^2 (Пірсона), а також показників описової статистики (середнього арифметичного вибірки й стандартного відхилення) при проведенні статистичного аналізу даних в дослідженнях.

Ключові слова: наукове дослідження, статистичний аналіз даних, критерій Пірсона, майбутні фахівці, MS Excel.

В наукових дослідженнях широко використовуються математичні методи, що є процедурами побудови, перетворення, метризації і обчислення математичних об'єктів. Ці процедури ґрунтуються на теоріях множин, графів, матриць, вірогідності і математичної статистики [1].

Метою будь-якого експерименту є емпіричне підтвердження гіпотези дослідження і/або справедливості теоретичних результатів, тобто обґрунтування того, що запропонована педагогічна дія (наприклад, нові зміст, форми, методи, засоби навчання і т. ін.) ефективніша (чи, можливо, навпаки – менш ефективна). Для цього, як мінімум, необхідно показати, що, будучи застосованим до того ж об'єкту (наприклад, до групи учнів), вона дає інші результати, чим застосування традиційних педагогічних дій.

Використання комп'ютера під час аналізу результатів експериментів є, безумовно, доцільним. Усі чотири запропоновані для використання для обробки результатів експерименту статистичні критерії (Крамера-Уелча, Вілкоксона-Манна-Уїтні, критерій однорідності χ^2 (Пірсона), Фішера) коректно реалізовані в професійних статистичних пакетах, серед яких можна виділити і рекомендувати такі найбільш поширені пакети статистичного аналізу як: Statistika, StatGraphics, SPSS [2]. Проте, згадані програми є ліцензійними, при цьому вони досить складні і вимагають значних тимчасових витрат для свого освоєння. Разом з цими

професійними статистичними пакетами деякі інструменти статистичного аналізу містить популярний табличний процесор Microsoft Excel, що входить в стандартний комплект пакету Microsoft Office, який встановлений практично на будь-якому персональному комп'ютері.

Однак, слід зазначити, що жоден з чотирьох рекомендованих статистичних критеріїв не реалізований в Excel, тому для автоматизації розрахунку емпіричних значень критеріїв слід ввести вручну в електронну таблицю формули, а також скористатися можливостями інструмента аналізу даних Excel „Описова статистика”. Таким чином, використання для проведення статистичного аналізу даних експерименту електронних таблиць Excel є актуальним.

Мета статті – використання табличного процесору Microsoft Excel для автоматизації розрахунку емпіричного значення критерію однорідності χ^2 , а також показників описової статистики при проведенні статистичного аналізу даних в дослідженнях.

Типовим завданням аналізу даних у дослідженнях є встановлення збігів або відмінностей характеристик експериментальної (ЕГ) та контрольної груп (КГ). Для цього формулюються статистичні гіпотези: нульова (гіпотеза про відсутність відмінностей) та альтернативна (гіпотеза про значущість відмінностей). Для прийняття рішення про певну гіпотезу використовують статистичні критерії (наприклад, критерій однорідності χ^2), тобто розраховується емпіричне значення критерію ($\chi^2_{емп.}$), яке потім порівнюється з табличним числом, яке називається критичним значенням критерію ($\chi^2_{критич.}$). У педагогічних дослідженнях зазвичай використовують рівень значущості 0,05 для критичного значення критерію.

На констатувальному етапі експерименту для достовірності досліджень потрібно перевірити нульову гіпотезу: відсутність суттєвої різниці, тобто

Таблиця 1

Усереднений розподіл студентів за рівнями сформованості готовності в ЕГ та КГ за всіма критеріями (до експерименту)

Група	Абс. кількість	Рівні сформованості готовності до використання КТ					
		низький		середній		високий	
		абс. кількість	(%)	абс. кількість	(%)	абс. кількість	(%)
ЕГ	118	54	45,8%	51	42,9%	13	11,3%
КГ	115	53	46,0%	48	42,2%	14	11,8%

збіг на заданому рівні значущості характеристики експериментальної та контрольної груп.

Усереднений розподіл студентів за рівнями сформованості готовності до використання комп'ютерних технологій (КТ) до використання у професійній діяльності за всіма критеріями на констатувальному етапі експерименту представлений у таблиці 1 і на рисунку 1.

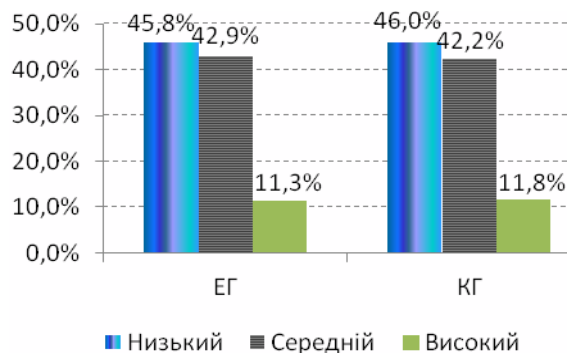


Рис. 1. Усереднений розподіл студентів за рівнями сформованості готовності у ЕГ та КГ за всіма критеріями (до експерименту)

Спираючись на кількісні показники, наведені в табл. 1, перевіримо нульову гіпотезу за формулою [2, с. 53 – 54]:

$$\chi^2_{\text{емп.}} = N \cdot M \cdot \sum_{i=1}^L \frac{(n_i - m_i)^2}{n_i + m_i},$$

де N – обсяг вибірки в експериментальній групі (кількість студентів);

M – обсяг вибірки в контрольній групі;

$n = (n_1, n_2, \dots, n_L)$, де n_i – число членів експериментальної групи, які одержали i -й бал, $i = 1, 2, \dots, L$;

$m = (m_1, m_2, \dots, m_L)$, де m_i – число членів контрольної групи, які одержали i -й бал, $i = 1, 2, \dots, L$;

L – кількість рівнів. У нашому випадку $L = 3$ – „низький”, „середній” або „високий” рівень.

За даними таблиці „Відсоткові точки розподілу χ^2 ”, для числа ступеня свободи (у нашому випадку це число 2) і рівня значущості 0,05, $\chi^2_{\text{критич.}} = 5,99$.

Використовуючи дані таблиці 1 та можливості електронних таблиць MS Excel щодо статистичного аналізу, ми підраховали значення критерію $\chi^2_{\text{емп.}}$ для експериментальних груп (рис. 2). Отримані дані представлено в таблиці 2.

Таблиця 2

Зіставлення емпіричного та критичного значень критерію Пірсона χ^2 (констатувальний зріз)

Емпіричне значення критерію ($\chi^2_{\text{емп.}}$)	Критичне значення критерію ($\chi^2_{\text{критич.}}$) ($v=2, p \leq 0,05$)
0,1	5,99

Порівняння емпіричного значення критерію з критичним значенням ($\chi^2_{\text{емп.}} = 0,1 \leq \chi^2_{\text{критич.}} = 5,99$) дозволяють стверджувати, що характеристики експериментальної й контрольної груп збігаються з рівнем значущості 0,05 за статистичним критерієм однорідності. Тобто початковий (до експерименту) стан готовності експериментальної та контрольної груп збігаються.

	ЕГ	КГ		$\chi^2_{\text{емп.}} = 0,1 < 5,99 = \chi^2_{\text{критич.}}$
Низький	54	53	9,82566E-08	Розбіжності не достовірні, тобто студенти ЕГ та КГ мають однаковий рівень сформованості готовності до використання КТ у
Середній	51	48	2,21614E-06	
Достатній	13	14	4,95765E-06	
Вибірка	118	115	7,27205E-06	
		$\chi^2_{\text{емп.}}$	0,10	
		$\chi^2_{\text{критич.}}$	5,99	

Рис. 2. Копія екрана ПК з розрахунками емпіричного значення критерію Пірсона χ^2 (MS Excel) до формувального експерименту для ЕГ та КГ

Також у статистичних критеріях для визначення достовірності збігів (або відмінностей) характеристик між експериментальною та контрольною групою використовуються деякі показники описової статистики [4]. Так, відсутність достовірних відмінностей між експериментальною та контрольною групами, які брали участь в експерименті, підтверджують близькі значення середнього арифметичного значення вибірки ($X=1,65$ – у ЕГ; $X=1,66$ – у КГ) і стандартного відхилення ($S=0,67$ – у ЕГ; $S=0,69$ – у КГ), які також

свідчать про схожість груп за рівнем сформованості готовності до використання КТ у професійній діяльності. Ці дані ми одержали за допомогою „Описової статистики” з надбудови „Аналіз даних” програми MS Excel (див. рис. 3).

ЕГ		КГ	
Среднее	1,652542373	Среднее	1,660869565
Стандартная ошибка	0,061815123	Стандартная ошибка	0,064044889
Медиана	2	Медиана	2
Мода	1	Мода	1
Стандартное отклонение	0,471484115	Стандартное отклонение	0,468804915
Дисперсия выборки	0,450890917	Дисперсия выборки	0,471700992
Эксцесс	-0,715827705	Эксцесс	-0,76324742
Асимметричность	0,543872129	Асимметричность	0,55647677
Интервал	2	Интервал	2
Минимум	1	Минимум	1
Максимум	3	Максимум	3
Сумма	195	Сумма	191
Счет	118	Счет	115
Уровень надежности(95,0%)	0,122421612	Уровень надежности(95,0%)	0,126872431

Рис. 3. Копія екрана ПК з розрахунками середнього арифметичного вибірки й стандартного відхилення (MS Excel) до формувального експерименту для контрольної та експериментальної груп

Підсумкові дані усередненого розподілу всіх студентів ЕГ та КГ за рівнями сформованості готовності до використання КТ за всіма критеріями на формувальному етапі експерименту наведено у таблиці 3.

Таблиця 3

Усереднений розподіл студентів ЕГ та КГ за рівнями сформованості готовності до використання КТ у професійній діяльності за всіма критеріями (після експерименту)

Група	Ви-бірка	Рівні сформованості готовності						X	S
		Низький		Середній		Високий			
		Кіл.	%	Кіл.	%	Кіл.	%		
ЕГ	118	24	20,6	66	56,1	28	23,3	2,03	0,67
КГ	115	51	44,5	50	43,3	14	12,3	1,68	0,68

Ми також підраховували, використовуючи дані табл. 3 та надбудову MS Excel „Аналіз даних”, значення показників описової статистики X та S (середнє арифметичне значення вибірки та стандартне відхилення) для експериментальної та контрольної груп (рис. 4).

На формувальному етапі експерименту для емпіричного підтвердження гіпотези дослідження про ефективність запроваджених педагогічних умов підготовки майбутніх фахівців до використання КТ у професійній діяльності слід прийняти статистичну гіпотезу про значущість відмінностей для підтвердження наявності відмінностей між експериментальною й контрольною групою після експерименту.

ЕГ		КГ	
Среднее	2,03089811	Среднее	1,67626087
Стандартная ошибка	0,061291317	Стандартная ошибка	0,063881702
Медиана	2	Медиана	2
Мода	2	Мода	1
Стандартное отклонение	0,46079692	Стандартное отклонение	0,462401768
Дисперсия выборки	0,4402853	Дисперсия выборки	0,461751210
Эксцесс	-0,70412182	Эксцесс	-0,77931051
Асимметричность	-0,61767539	Асимметричность	0,507757339
Интервал	2	Интервал	2
Минимум	1	Минимум	1
Максимум	3	Максимум	3
Сумма	240	Сумма	202
Счет	118	Счет	115
Уровень надежности(95,0%)	0,12118475	Уровень надежности(95,0%)	0,128069759

Рис. 4. Копія екрана ПК з розрахунками середнього арифметичного вибірки й стандартного відхилення (MS Excel) після формувального експерименту для контрольної та експериментальної груп

За даними таблиці „Відсоткові точки розподілу χ^2 ”, для числа ступеня свободи (у нашому випадку це число 2) і рівня значущості 0,05, $\chi^2_{критич.} = 5,99$.

Використовуючи дані таблиці 3 та можливості електронних таблиць MS Excel щодо статистичного аналізу, ми підраховували значення критерію $\chi^2_{емп.}$ для ЕГ та КГ, які брали участь у формувальному етапі експерименту (рис. 4). Отримані дані з урахуванням даних табл. 2 представлено в табл. 4.

Таблиця 4

Результати зіставлення емпіричного та критичного значень критерію Пірсона χ^2 для експериментальної та контрольної груп до й після експерименту

Емпіричне значення критерію, ($\chi^2_{емп.}$)		Критичне значення ($\chi^2_{критич.}$) ($v=2, p \leq 0,05$)
ЕГ та КГ до експерименту	ЕГ та КГ після експерименту	
0,1	16,28	5,99

Оскільки отримане емпіричне значення критерію Пірсона χ^2 після експерименту ($\chi^2_{емп.} = 16,28$) є суттєво більшим, ніж $\chi^2_{критич.}$, тому є підстави стверджувати про наявність відмінностей між експериментальною й контрольною групами після експерименту (див. рис. 5).

ЕГ		КГ		$\chi^2_{емп.} = 16,28 > 5,99 = \chi^2_{критич.}$
Низький	24	51	0,000751913	
Середній	66	50	0,000130534	
Достатній	28	14	0,000314433	
Вибірка	118	115	0,001195879	
	$\chi^2_{емп.}$		16,28	
		$\chi^2_{критич.}$	5,99	

Рис. 5. Копія екрана ПК з розрахунками емпіричного значення критерію Пірсона χ^2 (MS Excel) після формувального експерименту для ЕГ та КГ

Порівняльний аналіз результатів зіставлення емпіричного та критичного значень критерію Пірсона χ^2 для експериментальної та контрольної груп до й після експерименту (див. табл. 4) дозволив зробити висновок про позитивні результати проведеної роботи. Кількісні дані отримали підтвердження за допомогою методів математичної статистики. Так, початкові (до початку експерименту) стани експериментальної й контрольної груп збігаються, а кінцеві (після закінчення експерименту) – розрізняються.

Відповідно до правила прийняття рішення є достатні підстави вважати, що рівень готовності студентів до використання КТ у професійній діяльності в експериментальній групі на початок і закінчення формувального експерименту різний, що цілком узгоджується з висунутою гіпотезою про ефективність запроваджених умов підготовки майбутніх фахівців до використання комп'ютерних технологій у професійній діяльності.

Таким чином, використання табличного процесору MS Excel для розрахунку емпіричного значення критерію (у тому числі χ^2), середнього арифметичного вибірки й стандартного відхилення надає можливість автоматизувати проведення статистичного аналізу даних в дослідженнях.

Відповідно до перспектив подальшого дослідження необхідно розглянути проведення розрахунків щодо статистичного аналізу даних у табличному процесорі MS Excel з використанням інших рекомендованих статистичних критеріїв.

Література

1. Красильников В. В. Математические методы в психолого-педагогических исследованиях : учебно-метод. пособие / В. В. Красильников, В. С. Тоискин. – Ставрополь : Изд-во СГПИ, 2008. – 84 с.
2. Новиков Д. А. Статистические методы в педагогических исследованиях (типовые случаи) / Д. А. Новиков. – М. : МЗ-Пресс, 2004. – 67 с.
3. Крамаренко Т. А. Підготовка майбутніх інженерів педагогів до використання комп'ютерних технологій у професійній діяльності : автореф. дис. на здобуття наук. ступеня канд. пед. наук : спец. 13.00.04 „Теорія і методика професійного навчання” / Т. А. Крамаренко. – Луганськ, 2013. – 20 с.
4. Адаменко Е. В. Математические методы в педагогике и психологии : учеб. пособие / Е. В. Адаменко. – Луганск : Альма-матер, 2008. – 94 с.

References

1. Krasil'nikov V. V. Matematicheskie metody v psihologo-pedagogicheskikh issledovaniyah : uchebno-metod. posobie / V. V. Krasil'nikov, V. S. Toiskin. – Stavropol' : Izd-vo SGPI, 2008. – 84 s.
2. Novikov D. A. Statisticheskie metody v pedagogicheskikh issledovaniyah (tipovye sluchai) / D. A. Novikov. – M. : MZ-Press, 2004. – 67 s.
3. Kramarenko T. A. Pidgotovka majbutnih inzheneriv pedagogiv do vikoristannja komp'juternih tehnologij u profesijnij dijalnosti : avtoref. dis. na zdobuttja nauk. stupenja kand. ped. nauk : spec. 13.00.04 „Teorija i

metodika profesijnogo navchannja” / Т. А. Крамаренко. – Луганск, 2013. – 20 с.

4. Adamenko E. V. Matematicheskie metody v pedagogike i psihologii : ucheb. posobie / E. V. Adamenko. – Lugansk : Al'ma-mater, 2008. – 94 s.

Крамаренко Т. А. Автоматизация обработки данных статистического анализа в научных исследованиях

В статье рассмотрено использования табличного процессора Excel для автоматизации расчета эмпирического значения критерия однородности χ^2 (Пирсона), а также показателей описательной статистики (среднего арифметического выборки и стандартного отклонения) при проведении статистического анализа данных в педагогических исследованиях.

Ключевые слова: педагогическое исследование, статистический анализ данных, критерий Пирсона, будущие специалисты, MS Excel.

Kramarenko T. A. Automation of treatment of data of statistical analysis in researches

In the article the uses of tabular processor of Excel are considered for automation of calculation of empiric value of criterion of homogeneity χ^2 (Pearson), and also indexes of descriptive statistics (middle arithmetic selection and standard deviation) during realization of statistical analysis of data in researches.

Key words: pedagogical research, statistical analysis of data, criterion of Pearson, future professionals, MS Excel.

Крамаренко Т. А., кандидат педагогічних наук, доцент кафедри інформаційних технологій та систем, ДЗ „Луганський національний університет імені Тараса Шевченка”

Рецензент: **Носко П.Л.**, д.т.н., професор.

Стаття подана 15.03.2014