

Особливості використання табличного процесора *Microsoft Excel* для статистичного аналізу емпіричних даних

У статті розкриваються специфічні особливості, які потрібно враховувати при використанні табличного процесора *Microsoft Excel* у процесі статистичного аналізу емпіричних даних.

Ключові слова: емпіричні дані, статистичний аналіз, засоби *Microsoft Excel* для статистичного аналізу даних.

Постановка наукової проблеми та її значення. Невід'ємною складовою частиною будь-якого дослідження є статистичний аналіз емпіричних даних, оскільки його результати дають змогу здійснити наукове обґрунтування висунутих на початку дослідження гіпотез, встановити взаємозв'язки, виявити закономірності та тенденції в досліджуваних явищах і процесах. Цей процес передбачає використання математичних методів, зокрема методів математичної статистики, оскільки об'єкти соціально-педагогічних досліджень здебільшого відповідають визначенню поняття випадкових явищ. Застосування цих методів у соціально-педагогічних дослідженнях супроводжується значними труднощами, які зумовлені структурною та функціональною складністю і багатофакторністю соціально-педагогічних явищ й процесів, необхідністю оперувати предметами, які важко піддаються кількісному аналізу (поглядами, смаками, установками, ціннісними орієнтаціями тощо), розвиненістю і складністю понятійного та методичного апарату математичної статистики.

Значною мірою обійти ці труднощі дають змогу сучасні інструментальні засоби інформаційних технологій, такі як SPSS, STATISTICA, OCA та ін., проте вони вимагають від користувача відповідного рівня підготовки до роботи з ними. На нашу думку, більш продуктивним шляхом для дослідника, який не має у своєму розпорядженні вищеозначених програмних продуктів, або ж не має достатнього рівня підготовки для роботи з ними, є використання для статистичного аналізу даних табличних процесорів, наприклад *Microsoft Excel*. Цей інструментальний засіб є достатньо поширеним у середовищі користувачів і включає достатньо широкий арсенал засобів, призначений для статистичного аналізу емпіричних даних.

Аналіз досліджень цієї проблеми. Питанням використання табличного процесора *Microsoft Excel* у процесі статистичного аналізу емпіричних даних присвячено цілий ряд наукових досліджень, зокрема праці таких науковців, як: Е. В. Чекотковський, В. А. Кошеев, І. В. Лупан, О. В. Авраменко, О. А. Василенко, І. А. Сенча та ін. Водночас, застосовуючи цей програмний продукт для коректного математично-статистичного аналізу даних, слід зважати на певні особливості, на яких не зацентовано достатньо уваги в сучасних наукових дослідженнях.

Мета статті – розкрити особливості використання засобів табличного процесора *Microsoft Excel* для статистичного аналізу даних, отриманих у процесі емпіричних соціально-педагогічних досліджень.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Засоби *Microsoft Excel*, які дають змогу виконувати статистичний аналіз даних, можна розділити на три рівні [1; 4; 6]:

1. Формули користувача, які розміщуються в комірках електронної таблиці, й окремі формули та етапи статистичних розрахунків (обчислення середнього значення, розрахунок відхилень, квадратів відхилень, сум квадратів відхилень, добутків відхилень та їх суми тощо), які реалізують за допомогою елементарних математичних операцій та функцій.

2. Статистичні функції, які являють собою функціональні програмні модулі, що реалізують окремі статистичні формули (розрахунок середніх значень дисперсії, коефіцієнта кореляції, довірчого (надійного) інтервалу тощо), і можуть використовуватись у формулах. *Microsoft Excel* містить кілька десятків статистичних функцій, робота з якими ведеться за допомогою спеціальної програми майстра функцій.

3. Надбудова Аналіз даних (пакет аналізу), який являє собою пакет спеціалізованих програм (макрофункцій), призначених для розв'язання складних статистичних та інженерних завдань із поданням результатів у різних формах, включаючи графічну, і набір звітів.

Крім того, існує можливість створення власних програм обробки даних на основі *Excel* за допомогою мови програмування *Visual Basic for Applications (VBA)*, яка входить до системи *Office* як загальний внутрішній засіб програмування і розширення можливостей аплікацій.

Використовуючи ці засоби, слід враховувати певні особливості, зокрема:

1. При застосуванні функцій слід пам'ятати, що у формулах *Excel* передбачається вкладеність функцій, тобто аргументом функції може бути формула, яка містить інші функції. Використання багаторівневої вкладеності функцій дає можливість зробити розрахункові таблиці більш компактними, однак ускладнює їх розуміння, фіксацію помилок і корекцію. У цьому процесі слід враховувати той факт, що у формулах можна використовувати до семи рівнів вкладеності.

2. У науковій літературі, яка стосується математичної статистики [2; 5], при ранжуванні емпіричних даних використовують різні підходи:

- а) найбільшому значенню в масиві даних присвоюється найвищий ранг; б) найбільшому значенню в масиві даних присвоюється найнижчий ранг.

Тому, використовуючи *Microsoft Excel* для ранжування даних, потрібно враховувати таке:

- для реалізації підходу а) при ранжуванні емпіричних даних за допомогою функції РАНГ у *Microsoft Excel*, задаючи порядок, потрібно вказати довільне ненульове число, а для підходу б) – число 0, або не задавати порядку;
- для того, щоб при копіюванні формули для обчислення рангу в інші комірки не відбувалася автоматична модифікація адреси діапазону комірок, посилання на діапазон комірок (наприклад A1:A10), у яких містяться дані, що ранжуються, потрібно наводити у вигляді абсолютного посилання (\$A\$1:\$A\$10);
- функція РАНГ присвоює значення, які повторюються, однаковий ранг – число, яке відображає позицію цього значення в масиві всіх значень, відсортованому відповідним чином. Наприклад, якщо числа 11 у масиві емпіричних даних, відсортованому за спаданням, займають 5-ту і 6-ту позицію, то їм обом *Microsoft Excel* присвоїть ранг 5. У науковій літературі ранг таких значень обчислюється як середнє між рангами цих значень, якби вони були різними за величиною, тобто кожному числу має бути присвоєний ранг 5,5. Тому після ранжування даних процесором потрібно коригувати величину спряжених рангів.

3. Існують певні особливості, на які потрібно звертати увагу під час розрахунків точкових оцінок характеристик розподілу випадкової величини, таких як «дисперсія», «асиметрія», «ексцес» та ін. Проілюструємо їх на прикладі такої оцінки як «ексцес». Наприклад, у літературі з математичної статистики [2; 5] для обчислення значення ексцесу розподілу емпіричних даних наводиться така формула:

$$E = \frac{\sum (x_i - \bar{x})^4 / n}{s^4}$$

У *Microsoft Excel* алгоритм обчислення ексцесу програмно реалізований таким чином, що його значення обчислюється за формулою:

$$E = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \cdot \frac{\sum x_i^4 - 4\bar{x} \sum x_i^3 + 6\bar{x}^2 \sum x_i^2 - 4\bar{x}^3 \sum x_i}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

У загальному випадку відмінність між оцінками ексцесу, розрахованими за вищезначеними формулами, невелика (особливо за достатньо великого n) і може бути знехтувана, але вона може мати значення, коли математичне сподівання відоме. Зокрема, якщо вибіркове середнє не дорівнює математичному сподіванню, то потрібно враховувати поправку Бесселя.

4. Дуже часто для кращої візуалізації процесу обробки даних вони розміщуються компактно на робочому листі, в суміжних стовпцях. Застосування макروفункції «Описательная статистика» компоненту *Microsoft Excel* «Пакет аналізу» для пошуку узагальнюючих характеристик масиву емпіричних даних, які дають змогу судити про наявність або відсутність центральних тенденцій у досліджуваних масивах, оцінювати міру їх змінливості, приведе до того, що будуть виведені значення цих характеристик для даних, які розміщені в кожному стовпці робочого листа. Тому для отримання узагальнюючих характеристик усього масиву даних усі дані мають бути розміщені в одному стовпці робочого листа електронної таблиці.

5. При побудові розподілів частот використовується функція Частота, яка входить у перелік статистичних функцій пакета *Excel*. Ця функція задається як функція масиву, тобто її результат являє собою не окреме значення, а впорядковану групу чисел – масив, який повертається функцією в одній комірці, причому в ній відображається тільки одне значення. При її записі необхідно дотримуватися правил синтаксису, властивих функціям такого типу. Вони передбачають, що як аргументи функції використовують посилання на адреси комірок, що містять множину даних, для яких обчислюються частоти та посилання на адреси комірок, котрі містять розрядні інтервали групування. Крім того, на відміну від звичайних функцій, після вводу функції масиву для того, щоб масив було відображено у відповідному діапазоні комірок, потрібно, замість натискування клавіші ENTER, натиснути комбінацію клавіш CTRL+SHIFT+ENTER.

6. Для генерації масиву випадкових чисел у *Microsoft Excel* використовується функція СЛЧИС(). Слід пам'ятати, що ця функція задається без аргументів, якими у звичайних формулах виступають або конкретні дані, або адреси комірок робочого листа книги, де ці дані знаходяться. Крім того, за будь-якої зміни даних, котрі містяться в робочому листі, автоматично відбувається генерація нового масиву випадкових чисел функцією СЛЧИС(). Тому, коли на основі згенерованого масиву випадкових чисел мають виконуватися інші операції (наприклад, ранжування, сортування у випадковому порядку текстових чи числових даних), для збереження відповідності потрібно скопіювати в буфер обміну масив згенерованих випадкових чисел і вставити у вихідний діапазон їх як значення через режим спеціальної вставки.

7. При обробці числових даних у *Microsoft Excel* користувач має можливість задавати точність їх відображення в комірках електронної таблиці. Наприклад, якщо при форматуванні комірки чи діапазону комірок користувач указав число десяткових знаків 2, то число 5,287943 буде відображено як 5,28, хоча реально в пам'яті зберігається число в повному вигляді. Під час виконання математичних операцій на основі формул, які задає користувач, процесор ці дії проводить на основі величин чисел, які збережені в пам'яті, і в результаті може прослідковуватися невідповідність між величинами цих чисел, які відображаються в комірках, та результатами їх, наприклад додавання. Тому, коли передбачається отримання певних підсумкових значень на основі наявних даних, їх потрібно попередньо представити у відповідному вигляді, застосувавши до них, наприклад, функцію ОКРУГЛ.

Врахування цих та інших особливостей використання *Microsoft Excel* у процесі статистичного аналізу емпіричних даних дасть змогу досліднику уникнути багатьох проблем, які виникають у цьому процесі, та підвищити його ефективність.

Висновки і перспективи подальших досліджень. Отже, на сьогодні комп'ютерні засоби та відповідне програмне забезпечення (зокрема табличний процесор *Microsoft Excel*) зробили складний апарат математичної статистики доступним і простим у використанні для широкого кола спеціалістів різних предметних галузей. Для того, щоб за допомогою цього інструментарію ефективно досліджувати статистичний, нечіткий, неоднозначний характер залежностей і закономірностей, які реалізуються під дією великого числа випадкових факторів та виявляються тільки в масових явищах, користувач має враховувати особливості його використання.

Джерела та література

1. Василенко О. А. Математично-статистичні методи аналізу у прикладних дослідженнях : навч. посіб. / О. А. Василенко, І. А. Сенча. – Одеса : ОНАЗ ім. О. С. Попова, 2011. – 166 с.
2. Гласс Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стенли ; под общ. ред. Ю. П. Адлера. – М. : Прогресс, 1976. – 495 с.
3. Кошеев В. А. Автоматизация статистического анализа данных : пакеты прикладных программ / В. А. Кошеев. – М. : Наука, 1988. – 230 с.
4. Лупан І. В. Комп'ютерні статистичні пакети : навч.-метод. посіб / І. В. Лупан, О. В. Авраменко. – Кіровоград : [б. в.], 2010. – 218 с.
5. Сидоренко Е. В. Методы математической обработки в психологии / Е. В. Сидоренко. – СПб. : Речь, 1996. – 212 с.
6. Чекотковський Е. В. Графіки статистичних рядів та їх побудова на ПЕОМ з використанням пакета Excel / Е. В. Чекотковський. – К. : КНЕУ, 1997. – 380 с.

Гулько Степан. Особенности использования табличного процессора *Microsoft Excel* для статистического анализа эмпирических данных. В статье раскрываются специфические особенности, которые необходимо учитывать при использовании табличного процессора *Microsoft Excel* в процессе статистического

анализа эмпирических данных. Основное внимание акцентируется на ранжировании эмпирических данных, особенностях вычислительных операций с помощью формул, сформированных пользователем, проведении рас-чётов точечных оценок характеристик распределений случайных величин, использовании компонента *Micro-soft Excel* «Пакет анализа», многоуровнености вложенности функций, особенностях работы с функциями, результатом выполнения которых есть не отдельное значение, а упорядоченная группа чисел – массив, который возвращается функцией в соответствующем диапазоне ячеек электронной таблицы.

Ключевые слова: эмпирические данные, статистический анализ, средства *Microsoft Excel* для статисти-ческого анализа данных.

Gun'ko Stepan. Features of Use of the Tabular *Microsoft Excel* Processor for the Statistical Analysis of Empirical Data. In article specific features which need to be considered when using the tabular *Microsoft Excel* processor in the course of the statistical analysis of empirical data reveal. The main attention is focused on ranging of empirical data, features of computing operations with the help of the formulas created by the user, carrying out calculations of dot estimates of characteristics of distributions of random variables, uses of the *Microsoft Excel* «Analysis Package» component, a multilevel of functions, features of work with the functions which result of performance there is not a separate value, and the ordered group of numbers – the massif which comes b ack function in the corresponding range of cells of a spreadsheet.

Key words: empirical data, the statistical analysis, means of *Microsoft Excel* for the statistical analysis of data.

Стаття надійшла до
редколегії
24.05.2014 р.