

5. Nikolic B., Ivanovic M., Milošević D. (2017). A Comparative Study of Columnar Storage and Indexing Techniques in Relational Database Management Systems. *Journal of Computer Science and Technology*, 17(4), 407–416.

УДК 004.62

*Колосова К. К., здобувачка I курсу спеціальності 122 Комп'ютерні науки,
Огороднік М. О., здобувачка I курсу спеціальності 122 Комп'ютерні науки,
Ніколюк П. К., д-р фіз.-мат. наук, професор, професор кафедри інформаційних технологій*

ПОРІВНЯННЯ РІЗНИХ МЕТОДІВ ЗНАХОДЖЕННЯ ЗНАЧЕННЯ КОРЕЛЯЦІЇ МІЖ НАБОРАМИ ЗМІННИХ

Донецький національний університет імені Василя Стуса, м. Вінниця

В основі багатьох кількісних досліджень лежить встановлення зв'язку між даними, а саме знаходження кореляції для розуміння, як пов'язані дві чи більше змінних. Дослідники зауважують, що кореляція не вказує на причинно-наслідкові зв'язки, особливо, коли цей критерій застосовується до перехресних даних. Припущення про вимірювання елементів тільки однією факторною величиною призводить до можливого завищення величини кореляції між змінними. Це обмеження викликало появу методів кореляції другого покоління, заснованих на моделюванні структурних рівнянь, як-от підтверджувальний факторний аналіз та дослідницьке моделювання структурних рівнянь. Точна оцінка кореляції між змінними має вирішальне значення для достовірності результатів дослідження. Коли дослідники отримують точні результати кореляції, це гарантує, що зв'язки між змінними представлені якомога ближче до реальності. Ця точність важлива, оскільки вона дає змогу дослідникам робити обґрунтовані та значимі висновки на основі своїх даних.

Якщо результати кореляції неточні або необ'єктивні, це може призвести до неправильної інтерпретації зв'язків між змінними. Наприклад, якщо кореляція завищена, це може свідчити про сильніший зв'язок між змінними, ніж існує насправді, що є результатом неправильних висновків і потенційно помилкових втручань, заснованих на помилкових доказах. У кількісних дослідженнях у сферах освіти та психології дослідники використовують різні методи для оцінки кореляції між змінними. Деякі приклади цих методів містять двофакторну кореляцію, підтверджувальний факторний аналіз та моделювання на основі дослідницьких структурних рівнянь.

Двофакторна кореляція передбачає обчислення коефіцієнта кореляції та є основним методом, який використовується для оцінки зв'язку між двома змінними (рис. 1). Цей показник вимірює силу та напрям лінійного зв'язку між наборами змінних. Він обчислює коефіцієнт кореляції, зазвичай коефіцієнт кореляції Пірсона, який вимірює силу та напрям лінійного зв'язку між двома змінними. Двофакторна кореляція передбачає, що змінні неперервні та нормально розподілені [1].

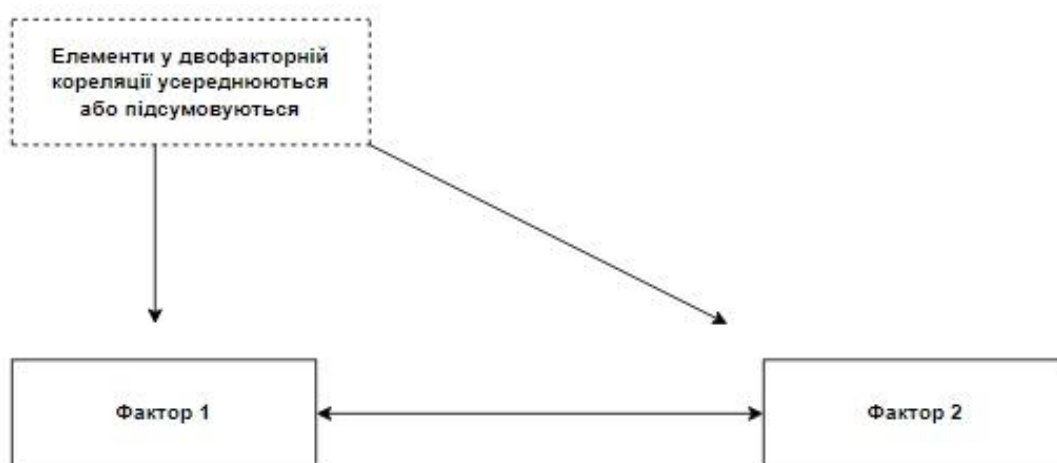


Рисунок 1. Візуальне представлення кореляції у методі двовимірної кореляції

Підтверджувальний факторний аналіз – це метод моделювання структурних рівнянь, що досліджує зв'язок між спостережуваними змінними та прихованими факторами (рис. 2). Його можна використовувати для оцінки кореляції між латентними змінними шляхом оцінки факторних навантажень і коваріацій. Цей статистичний метод використовується для перевірки моделі вимірювання й оцінки зв'язків між спостережуваними змінними та латентними конструктами. Він часто використовується в моделюванні структурними рівняннями [2].

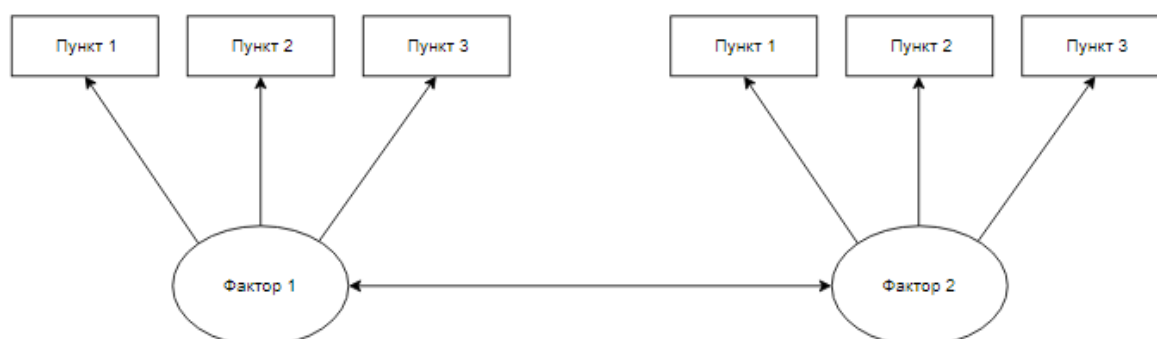


Рисунок 2. Візуальне представлення кореляції у методі підтверджувального факторного аналізу

Моделювання на основі дослідницьких структурних рівнянь – це інший тип моделювання на основі структурних рівнянь, що поєднує елементи дослідницького факторного аналізу та підтверджувального факторного аналізу (рис. 3). Цей метод дає змогу досліджувати як спостережувані, так і латентні змінні та може забезпечити більш точні оцінки кореляції, враховуючи складність і багатовимірність конструкцій. На відміну від підтверджувального факторного аналізу, метод моделювання на основі дослідницьких структурних рівнянь не припускає, що спостережувані змінні пов'язані виключно з їх гіпотетичними латентними конструктами. Це допускає можливість перехресних завантажень, коли спостережувані змінні можуть завантажуватися на кілька прихованих конструкцій. Цей метод забезпечує більш гнучку та точну оцінку факторних кореляцій, порівняно з підтверджувальним факторним аналізом [3].

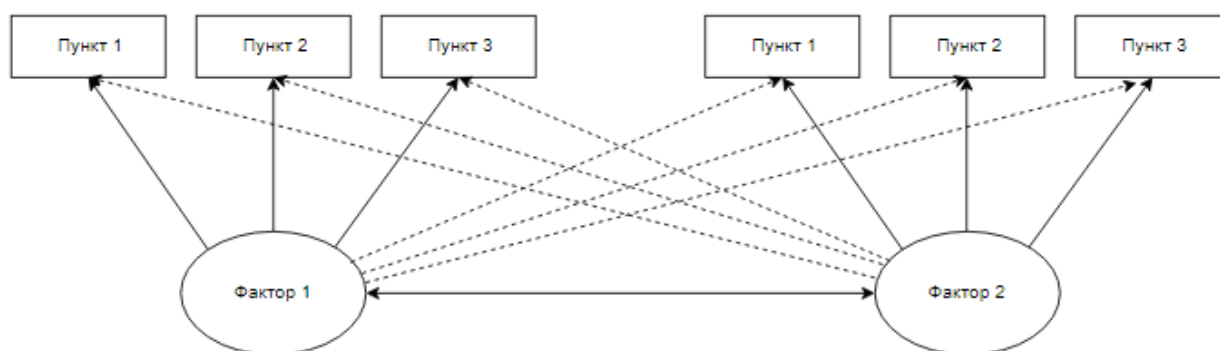


Рисунок 3. Візуальне представлення кореляції у методі моделювання на основі дослідницьких структурних рівнянь

Описані методи знаходження кореляції відрізняються своїми припущеннями та підходами до оцінки кореляції між змінними. Двовимірна кореляція є простим і зрозумілим методом, водночас підтверджувальний факторний аналіз та метод моделювання на основі дослідницьких структурних рівнянь забезпечують більш складні методи аналізу складних зв'язків між спостережуваними змінними і латентними конструктами.

Проаналізувавши методи знаходження кореляції у цій роботі, можна зробити висновки про те, що різні методи аналізу можуть давати різні результати та інтерпретації під час оцінки кореляції між змінними в кількісних дослідженнях. Можна виокремити обмеження методу двовимірної кореляції, що покладається на усереднення або підсумовування балів окремих елементів для представлення конструкцій. Цей метод може призвести до завищених кореляцій і перешкодити точній оцінці зв'язку між змінними. У дослідженнях рекомендують використовувати методи підтверджуючого факторного аналізу і дослідницького моделювання структурних рівнянь, які враховують багатовимірність конструкцій і допомагають точніше оцінювати кореляції між змінними. Загалом не можна

заперечувати важливість вибору відповідних методів для оцінки кореляції та необхідність використання більш точних методів, як-от методи підтверджуючого факторного аналізу і дослідницького моделювання структурних рівнянь.

Список використаних джерел

1. Hair, J. F., Hult, T., Ringle, C. M., Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 3rd edn. New York, NA: Sage.
2. Morin, S., Arens, K., Tran, A., Caci, H. (2016). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: a tutorial and illustration using the composite scale of morningness. *Int. J. Methods Psychiatr. Res.* 25, 277–288. DOI: 10.1002/mpr.1485.
3. Alamer, A., Marsh, H. (2022). Exploratory structural equation modeling in second language research: an applied example using the dualistic model of passion. *Stud. Second Lang. Acquis.* 1–24. DOI: 10.1017/S0272263121000863.

УДК 004.6+005.7

Крохмалюк В. В., здобувач 2 курсу спеціальності 122 Комп'ютерні науки ОС «Магістр»,

Потапова Н. А., канд. екон. наук, доцент, доцент кафедри інформаційних технологій

ВИКОРИСТАННЯ ІНСТРУМЕНТІВ ДЛЯ ЗБОРУ ДАНИХ У ДОСЛІДНИЦЬКІЙ ДІЯЛЬНОСТІ ЯК ЕФЕКТИВНИЙ МЕТОД ОТРИМАННЯ ІНФОРМАЦІЇ

Донецький національний університет імені Василя Стуса, м. Вінниця

Парсинг – це процес аналізу вхідних даних з метою виділення з них значимих елементів та перетворення їх у структурований формат для подальшої обробки. У програмуванні парсинг часто використовується для обробки текстових даних, як-от дані у форматі CSV, JSON або XML, а також для обробки коду мов програмування [2].

Парсинг може виконуватись із допомогою спеціальних програм, які називають парсерами. Парсер зчитує вхідні дані, аналізує їх за заданою граматиною та перетворює їх у структурований формат. Із допомогою парсера можна виділити з великого обсягу текстової інформації лише ту, що потрібна для подальшої обробки. У веброзробці парсинг часто використовується для збору даних із вебсторінок. Основна його ідея полягає в автоматичному зборі даних із вебсторінок за допомогою скриптів [1].

Парсинг може бути використаний для збору інформації про товари, послуги, ціни, новини тощо. Отже, парсинг – це процес аналізу вхідних даних з метою їх