3. System of Environmental-Economic Accounting 2012. Central Framework. United Nations. (2014). *unstats.un.org.* Retrieved from https://unstats.un.org/unsd/envaccounting/seeaRev/SEEA_CF_Final_en.pdf

4. The Regulation (EU) No 691/2011 of the European Parliament and of the Council of 6 July 2011 on European environmental economic accounts. Amended by the Regulation (EU) No 538/2014 of the European Parliament and of the Council of 16 April 2014. *eur-lex.europa.eu.* Retrieved from https://eur-lex.europa.eu/eli/reg/2011/691/2014-06-16

5. System of Environmental-Economic Accounting (SEEA). United Nations. *seea.un.org*. Retrieved from https://seea.un.org/

6. Україна 2030Е – країна з розвинутою цифровою економікою / Фіщук В. та ін.; Український інститут майбутнього. URL: https://strategy.uifuture.org/kraina-z-rozvinutoyu-cifrovoyu-ekonomikoyu.html

7. Про державну статистику: Закон України від 17.09.1992 р. № 2614-XII, станом на 19.04.2014 р. URL: https://zakon.rada.gov.ua/laws/show/2614-12

НОВІ ТЕНДЕНЦІЇ В ДОКАЗОВІЙ СТАТИСТИЦІ: ПРОБЛЕМИ ІМПУТАЦІЇ ДАНИХ

Ковтун Наталія Василівна,
доктор економічних наук, професор,
завідувач кафедри статистики та демографії,
Київський національний університет імені Тараса Шевченка;
Фаталієва Анастасія-Нарміна,
Statistician,
Data Management and Biometrics / Statistical Programming, Edetek

Проблема пропусків даних нерозривно пов’язана з якістю результатів статистичного спостереження. Один із способів вирішення цієї проблеми полягає в тому, щоб повністю виключити записи, які мають пропуски, що призводить до зменшення вибірки, а це, своєю чергою, впливає на достовірність результатів. Водночас неправильне використання відсутніх даних також може вплинути на довірчі інтервали. Наявність пропусків у даних, так само, як і аналіз тільки повних спостережень (після виключення спостережень з пропусками), може призвести до отримання зміщених результатів, і як наслідок – до спотворення висновків, які можуть бути зроблені за результатами дослідження, і прийняття неправильних рішень. До основних причин пропусків належать [1]:

- Виключення суб’єкта (об’єкта) з дослідження внаслідок невідповідності певним вимогам спостереження.
- Наявність несприятливої події.
- Відсутність результату у суб’єкта.
- Відсутність реєстрації.
- Помилки дослідників.

Можна визначити такі межі пропусків даних: 1) менше 5% пропусків вважається несуттєвим і не впливає на результати дослідження; 2) втрати 20% і більше даних ставлять під сумнів результати дослідження [2]. Отже, чим більша частка відсутніх даних, тим менша надійність висновків і тим складніше довести достовірність результатів дослідження. Як наслідок відсутні значення є потенційним джерелом упередження в ході дослідження. Виключення суб'єктів може впливати на сумісність груп та підгруп, що, своєю чергою, призводить до упередженості при оцінці результату, а також до зсуву оцінок.

Існують різні способи вирішення проблем пропусків даних. Найпростіший з них – це виключення суб'єкта з розрахунків. Але наслідками такого підходу є зниження обсягу вибірки, зменшення суттєвості статистичних висновків, зміна довірчого інтервалу (наприклад, звуження внаслідок недооцінки варіативності). Саме тому важливим при роботі з пропущеними даними є ідентифікація характеру пропусків, які можуть бути абсолютно випадковими (MCAR), пропуски випадкового порядку (MAR) і невипадкові пропуски. Це вимагає застосування відповідних методів оброки даних з пропущеними значеннями: виключення, заповнення, зважування і моделювання [3]. Усі ці методи дають різні результати за різних обсягів і характерах пропусків.

Нами була зроблена спроба оцінити результати застосування різних методів імпутації на прикладі вибірки, для якої були симульовані різні варіанти пропусків даних. Так, при 10% пропусків MCAR оцінки параметрів і *p*-value для двох факторів, отриманих в результаті застосування першої групи методів, були наближені до результатів, отриманих на повних даних. Середньоквадратичні похибки, розраховані за методом безумовного середнього і методом заповнення пропусків з послідовним підбором, були близькими до ідеалу, всі інші методи завищували оцінку. Найбільш схожим з вихідними даними був коефіцієнт детермінації при застосуванні ***методу заповнення пропусків з послідовним підбором***.

При 25% пропусків MCAR для коефіцієнта детермінації найменша оцінка була при застосуванні методу заповнення безумовним середнім значенням, а переоцінка – методу заповнення пропусків з послідовним підбором. За іншими підходами зміна була мінімальною. Отже, отримані оцінки параметрів і *p*-value в результаті застосування методу аналізу наявних випадків були більш наближені до значень, отриманих при побудові ***регресії на повних даних***.

При 50% пропусків MCAR фактор ваги став незначним при застосуванні методу аналізу повних спостережень. Найбільш точна оцінка змінної була отримана в результаті методу ***умовного середнього***. Найменш схожа – за ***послідовним підбором***. Проте можна виділити метод заповнення безумовним середнім, результати застосування якого були найбільш наближені до первинних даних.

За результатами імпутації 10% і 50% пропущених даних MAR- зміна в оцінці параметрів для вільного члена і двох інших факторів була мінімальною. При застосуванні методу множинної імпутації середньоквадра-

тична похибка і коефіцієнт детермінації були максимально близькі до результатів, отриманих на основі повних даних.

Проведене дослідження надало можливість виявити слабкі та сильні сторони кожного методу імпутації даних та визначити ефективність застосування того чи іншого методу при різних частках пропущеної інформації. Але безперечним результатом можна вважати доведення факту, що до процесу імпутації слід підходити дуже обережно, проблема імпутації має вирішуватися в кожному конкретному випадку на основі ретельного аналізу існуючої бази даних, з урахуванням особливостей не тільки самих даних і обсягу пропусків, а й конкретного дослідження.

Список використаних джерел

1. Little, R. J., D'Agostino, R., Cohen, M. L., & Dickersin, K. (2012). The Prevention and Treatment of Missing Data in Clinical Trials. *The New England Journal of Medicine [online]*. Retrieved from http://www.nejm.org/doi/pdf/10.1056/nejmsr1203730

2. Dziura, J. D., Post L. A., Zhao, Q., Fu, Z., & Peduzzi, P. (2013). Strategies for dealing with missing data in clinical trials: from design to analysis. Yale Journal of Biology and Medicine [online]. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3767219/

3. Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, Vol. 8, No. 1,* 3–15.

ДО ПИТАННЯ СУЧАСНИХ ПОГЛЯДІВ НА ЦИФРОВУ ЕКОНОМІКУ В УКРАЇНІ

Козлов Валерій Володимирович,
кандидат технічних наук, доцент,
доцент кафедри економіко-математичних дисциплін
та інформаційних технологій;
Томашевська Тетяна Володимирівна,
кандидат технічних наук, доцент,
доцент кафедри економіко-математичних дисциплін
та інформаційних технологій;
Національна академія статистики, обліку та аудиту

Сьогодні цифровізація - головний перспективний напрям розвитку української економіки та суспільства. Але її ефективність залежить від чіткості постановки задачі - місії, візії нашої взаємодії. Правильно організувати цей процес - великий крок до успіху [1].

Цифровий розвиток економіки – це комплекс механізмів, мотивацій розвитку цифрових технологій, а відтак, розгалуженої цифрової інфраструктури, задля використання можливостей держави, посилення її конкурентоздатності, зростання благополуччя громадян. Концепція розвитку цифрової економіки передбачає виконання низки заходів, мета яких -