

МЕТОДИ ЗМЕНШЕННЯ ДИСПЕРСІЇ В ОНЛАЙН ЕКСПЕРИМЕНТАХ

Виконала:

студентка 4-го року навчання
освітньої програми «Прикладна математика»,
Столярук Юлія Олександрівна

Керівник:

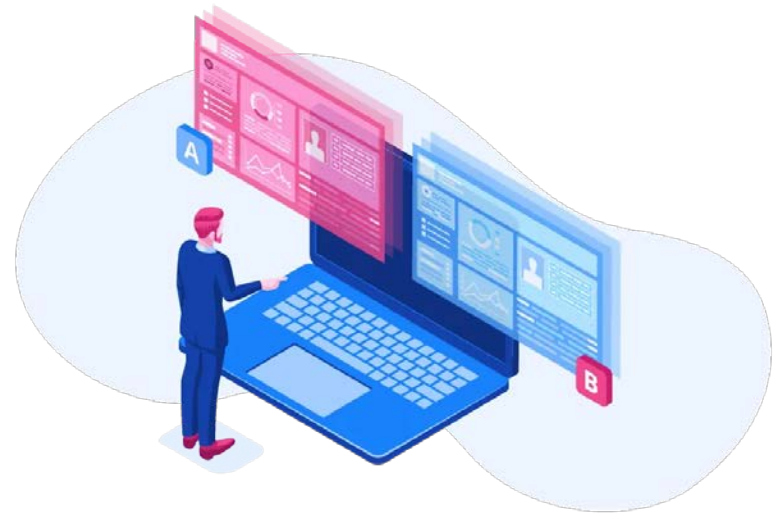
Крюкова Г. В.
Доцент, кандидат фізико-математичних наук



Мета дослідження

Мета дослідження: реалізувати та порівняти методи зменшення дисперсії оцінки ефекту втручання (ATE) в умовах A/A тестування.

Додатково: проаналізовано вплив кореляції між змінними та вибору експериментальної схеми на ефективність методів.





Основні поняття та теоретичні основи

Онлайн контрольований експеримент (англ. Online Controlled Experiment, або OCE) — це процедура, яка дозволяє встановити причинно-наслідковий зв'язок між зміною в цифровому продукті та поведінкою користувача.

- A/A тест ($K = 1$)
- A/B тест ($K = 2$)
- A/B/n тест ($K > 2$)
- K - кількість груп.
- Y_i - цільовий показника
- $Y_i(1)$ - показник у контрольній групі
- $Y_i(0)$ - показник в експериментальній групі



Середній ефект втручання (ATE)

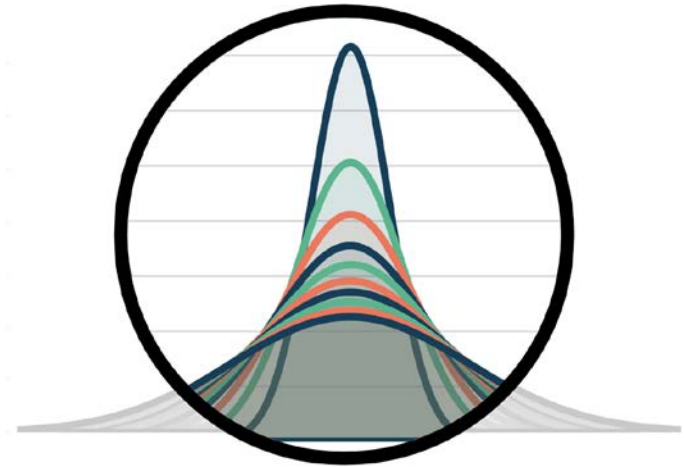
Середній ефект втручання (англ. average treatment effect, ATE) — це математичне сподівання різниці між потенційними результатами метрики Y за умов втручання ($T = 1$) та незмінного контролю ($T = 0$):

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)] = \mu(1) - \mu(0),$$

Методи зменшення дисперсії у онлайн експериментах дозволяють підвищити чутливість досліджень та скоротити розмір вибірки, зберігаючи точність результату.

Методи зменшення дисперсії:

- Стратифікація
- CUPED
- MLRATE
- STATE





Метод стратифікації

Метод зменшення дисперсії, що полягає в попередньому поділі експериментальних одиниць на підгрупи (страти) згідно з характеристиками, які суттєво впливають на результат.

$$\text{Var}(\hat{Y}_{\text{strat}}) = \frac{1}{n} \sum_{k=1}^K p_k \sigma_k^2 \quad (1)$$

$$\text{ATE}_{\text{strat}} = \sum_{k=1}^K p_k (\bar{Y}_k(1) - \bar{Y}_k(0)) \quad (2)$$



CUPED

CUPED (Controlled Using Pre-Experiment Data) - приклад методу коваріаційного коригування. Використовує допоміжну змінну X , яка фіксується до початку експерименту та є статистично залежною з основною метрикою Y .

$$Y_i^{\text{cuped}} = Y_i - \theta X_i + \theta \mathbb{E}[X] \quad (1)$$

$$\text{Var}(Y^{\text{cuped}}) = \text{Var}(Y)(1 - \rho^2) \quad (2)$$

$$\text{ATE}_{\text{cuped}} = \overline{Y^{\text{cuped}}}_{(1)} - \overline{Y^{\text{cuped}}}_{(0)} \quad (3)$$



MLRATE

MLRATE (Machine Learning Regression-Adjusted Treatment Effect) - поєднує коваріаційне коригування з прогнозуванням результативної змінної Y за допомогою моделей машинного навчання.

$$\hat{g}_i := \hat{g}_{k(i)}(X_i), \quad \bar{g} := \frac{1}{N} \sum_{i=1}^N \hat{g}_i. \quad (1)$$

$$Y_i = \alpha_0 + \alpha_1 T_i + \alpha_2 \hat{g}_i + \alpha_3 T_i (\hat{g}_i - \bar{g}) + \varepsilon_i \quad (2)$$

$$ATE_{MLRATE} = \alpha_1 \quad (3)$$



STATE

STATE (Student's t-distribution average treatment estimator) — оцінка середнього ефекту втручання (ATE), що поєднує машинне навчання та властивості розподілу Стьюдента.

$$\hat{g}_i := \hat{g}_{k(i)}(X_i), \quad \bar{g} := \frac{1}{N} \sum_{i=1}^N \hat{g}_i \quad (1)$$

$$Y_i = a_0 + a_1 T_i + a_2 \hat{Y}_i + \varepsilon_i \quad (2)$$

$$\text{ATE}_{STATE} = \alpha_1 \quad (3)$$

E-крок:

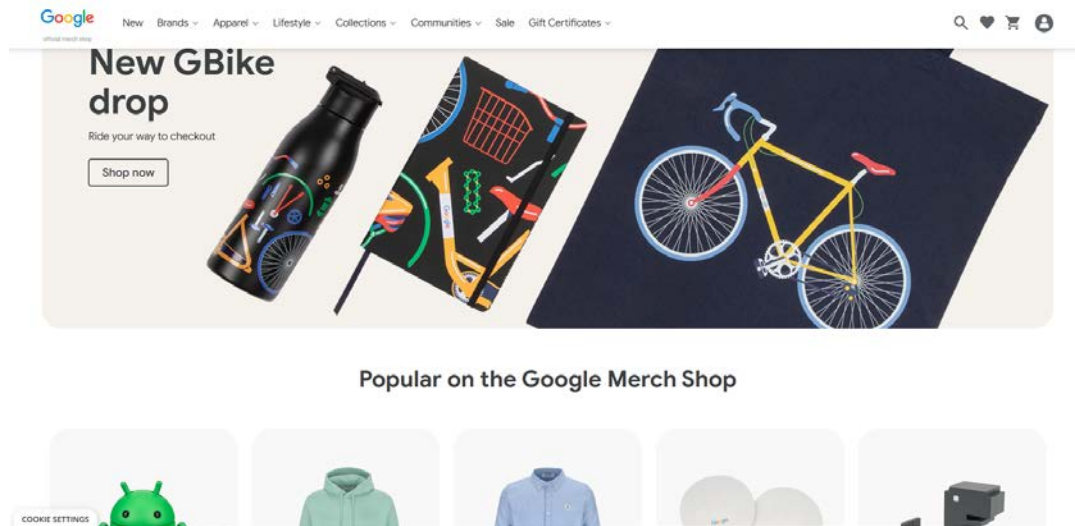
$$w_i^{(k)} = \frac{\nu + 1}{\nu + \frac{(Y_i - a_0^{(k)} - a_1^{(k)} T_i - a_2^{(k)} \hat{Y}_i)^2}{\sigma^{2(k)}}}$$

M-крок:

$$a^{(k+1)} = \arg \min_a \sum_i w_i^{(k)} (Y_i - a_0 - a_1 T_i - a_2 \hat{Y}_i)^2$$



Дані



- Джерело: Google (Google Merchant Store)
- Специфіка: метрики користувачів
- Період: 01.08.2016 - 01.08.2017
- Реалізація: Python

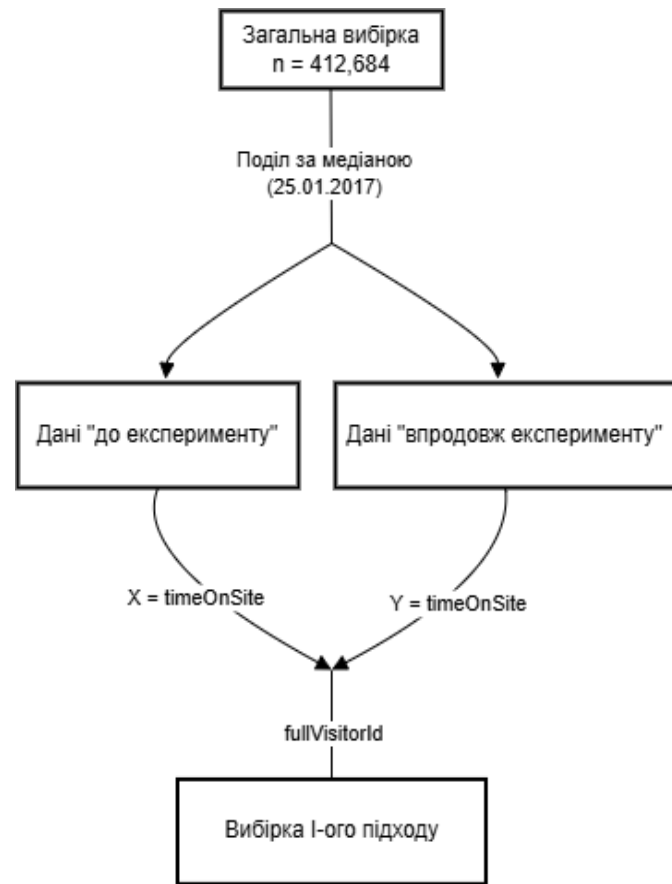


Постановка експерименту

- Обрано метрику *timeOnSite*— найменше пропущених значень.
- Немає явного розділення на контроль/експеримент -> моделювання A/A тесту.
- Реалізовано два підходи до побудови експерименту:
 1. доекспериментальні vs умовно експериментальні дані.
 2. усі дані умовно експериментальні використано корельовану змінну X.

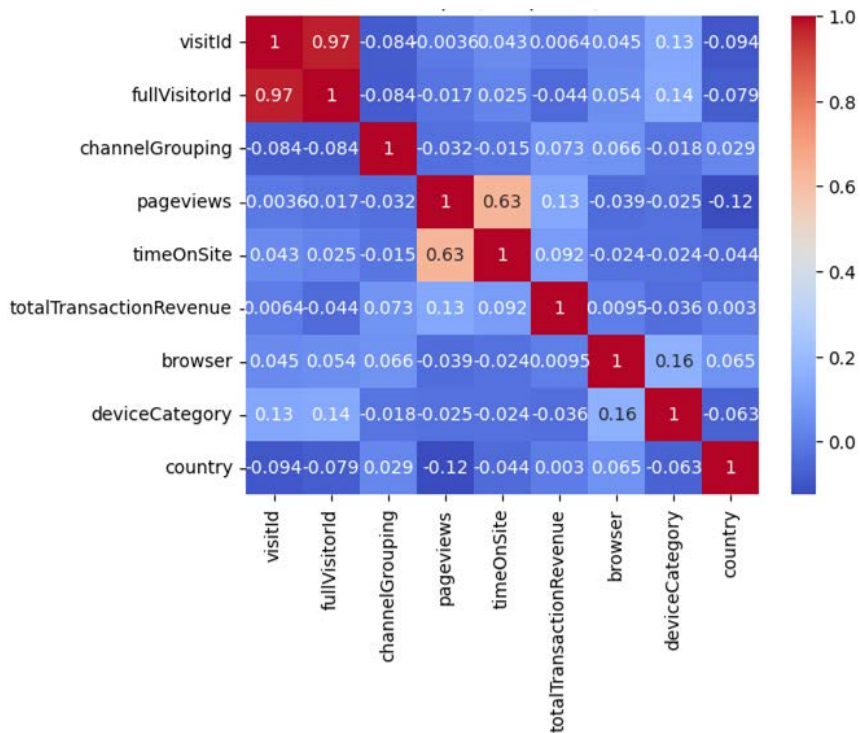
Перший спосіб

- Поділ на до- та післяекспериментальні спостереження за медіаною дати візиту (25.01.2017).
- Об'єднання даних за *fullVisitorId*, залишено лише тих користувачів, які мають спостереження в обох періодах.
- $n = 4,220$
- $X = \text{timeOnSite}$ до експерименту
- $Y = \text{timeOnSite}$ після
- $\text{cor}(X, Y) = 0.096$
- Підвибірki: по 100 записів, з випадковим поділом на $T = 0 / T = 1$



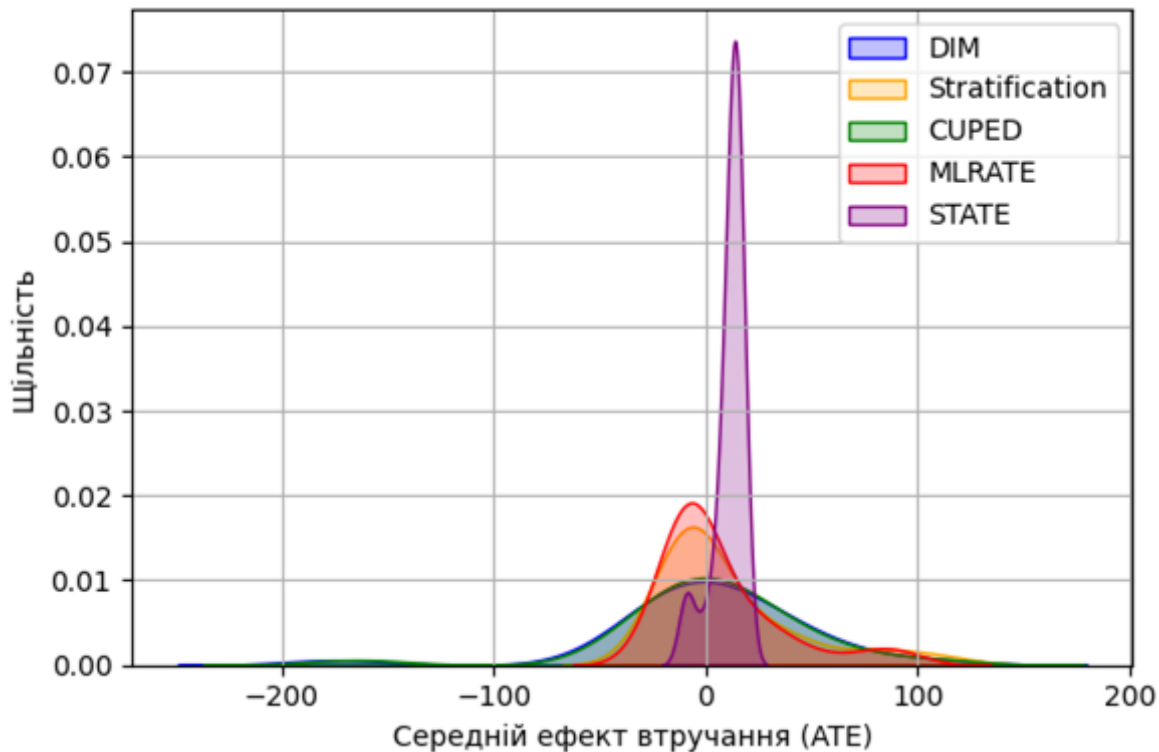


Другий спосіб

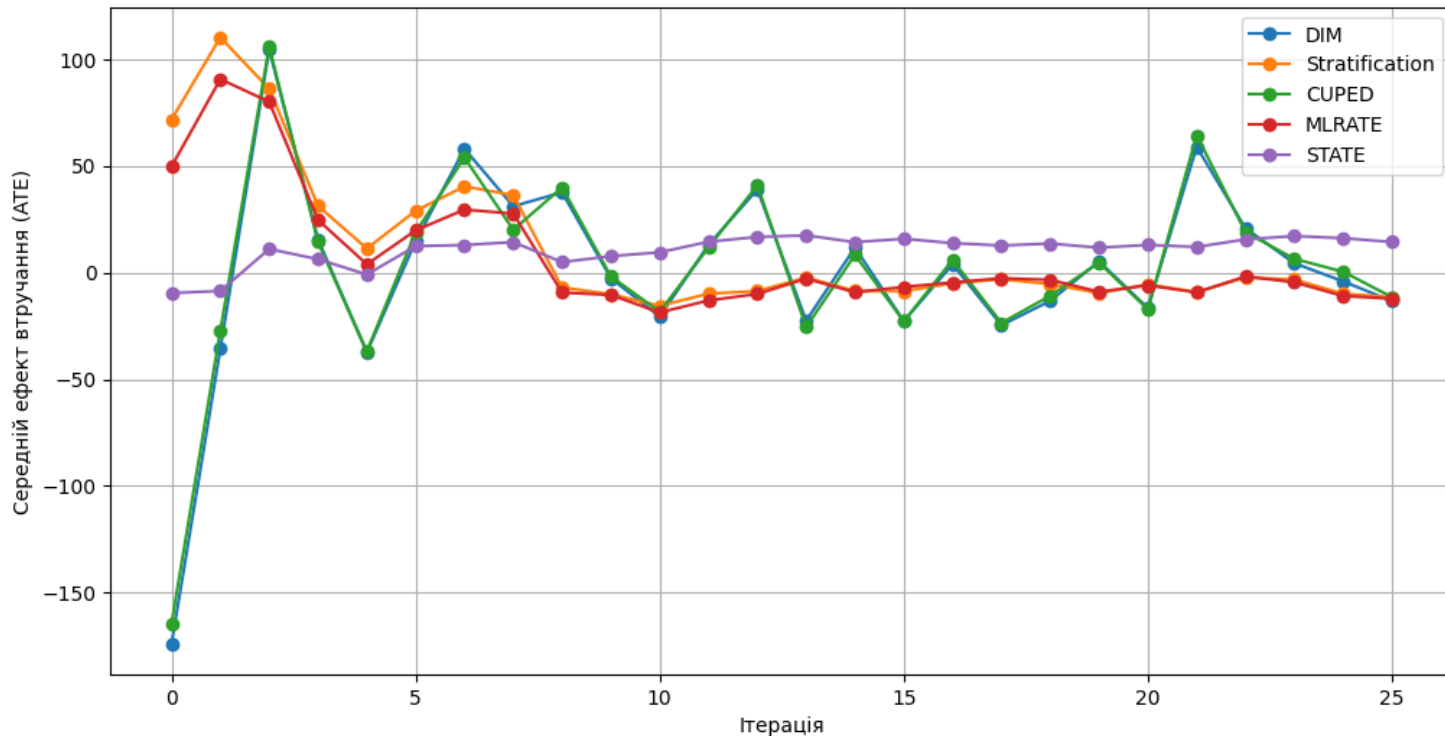


- Дані не мають чіткого розподілу на до- та післяекспериментальні періоди
- $X = \text{pageViews}$
- $Y = \text{timeOnSite}$
- $cor(X, Y) = 0.63$ (див. рис.)
- Обидві змінні належать до однієї сесії
- Використано всю вибірку : $n = 412,684$
- Підвибірки: по 10,000 записів, з випадковим поділом на $T = 0 / T = 1$

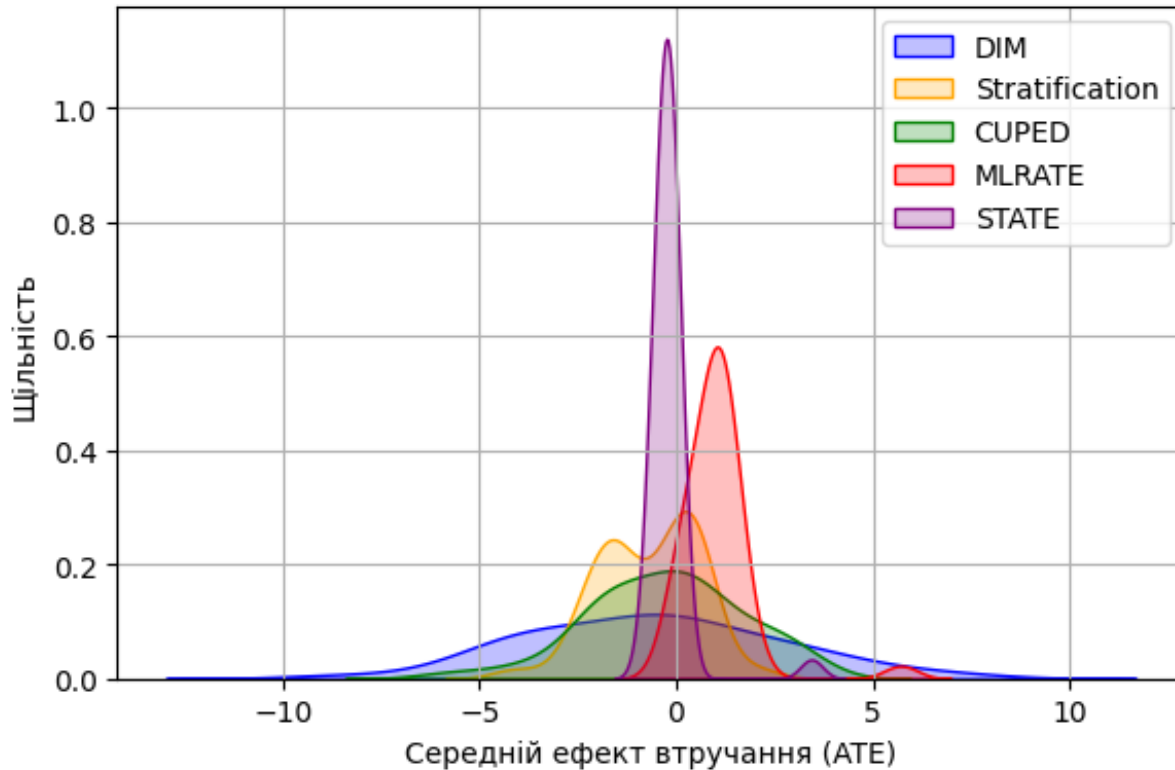
Щільність ефекту втручання (1-ий підхід)



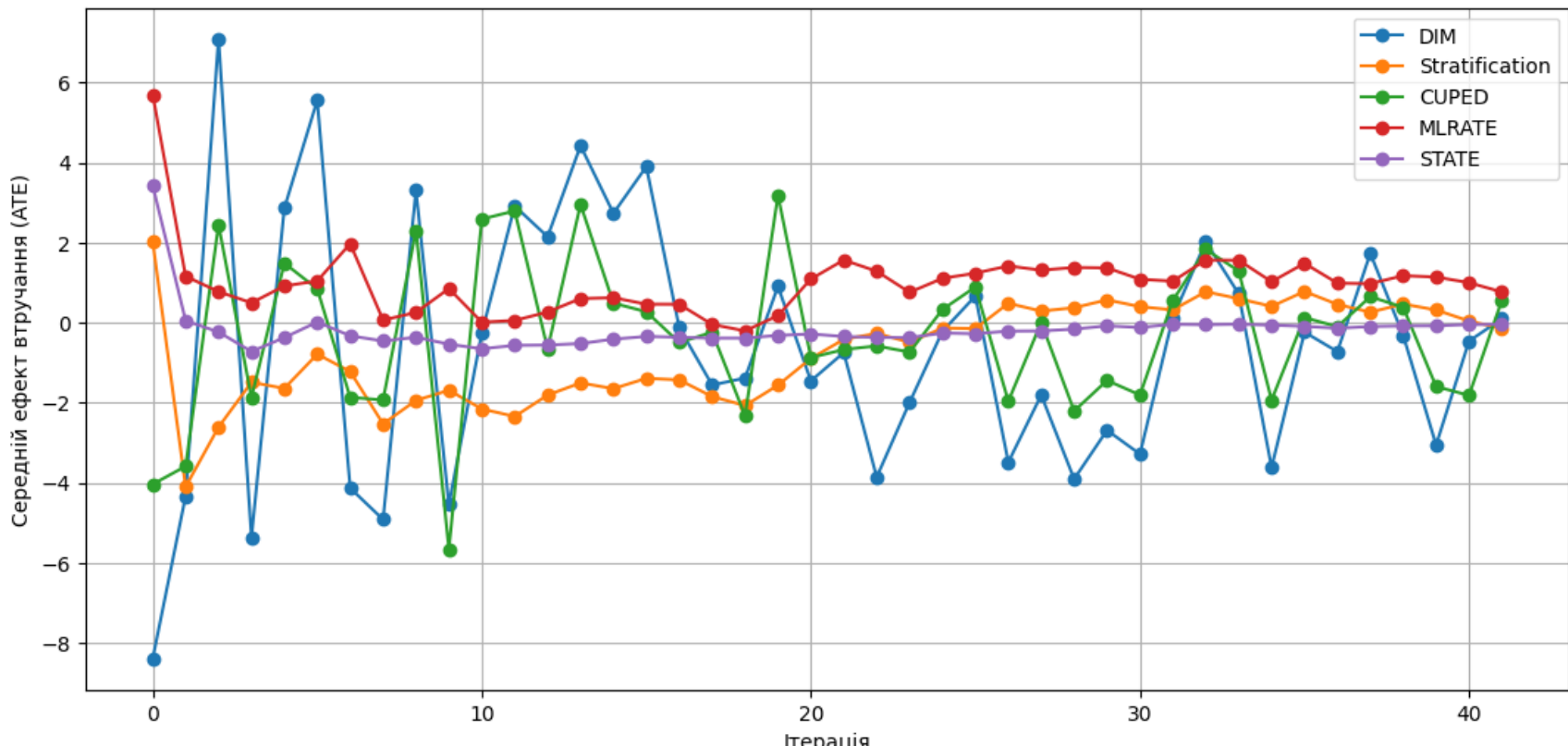
Вплив розміру вибірки (1-ий підхід)



Щільність ефекту втручання (2-ий підхід)



Вплив розміру вибірки (2-ий підхід)





Фінальні результати

Метод	Емпіричне покриття	Зменшення дисперсії
Стратифікація	100%	52.48%
CUPED	92.3%	6.34%
MLRATE	100%	65.87%
STATE	100%	97.82%

Табл. 1: Результати за першим підходом

Метод	Емпіричне покриття	Зменшення дисперсії
Стратифікація	100%	85.00%
CUPED	95.2%	63.64%
MLRATE	100%	92.18%
STATE	100%	96.53%

Табл. 2: Результати за другим підходом

Найкращі
результати
показує метод
STATE

Недолік цього
методу -
складність.



Література

1. Larsen N., Stallrich J., Sengupta S., Deng A., Kohavi R., Stevens N.T. “*Statistical Challenges in Online Controlled Experiments: A Review of A/B Testing Methodology.*” — The American Statistician, 2023.
2. Kohavi R., Tang D., Xu Y. “*Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing.*” — Cambridge University Press, 2020.
3. Kohavi R., Longbotham R., Sommerfield D., Randal H. “*Controlled experiments on the web: Survey and practical guide.*” — Data Mining and Knowledge Discovery, 2009.
4. Jin Y., Ba S. “*Towards Optimal Variance Reduction in Online Controlled Experiments.*” — Department of Statistics, Stanford University; Data Science Applied Research, LinkedIn Corporation, 2022.
5. Xie H., Aurisset J. “*Improving the Sensitivity of Online Controlled Experiments: Case Studies at Netflix.*” — 2016.
6. Deng A., Xu Y., Kohavi R., Walker T. “*Improving the Sensitivity of Online Controlled Experiments by Utilizing Pre-Experiment Data.*” — 2023.
7. Guo Y., Coey D., Konutgan M., Li W., Schoener C., Goldman M. “*Machine Learning for Variance Reduction in Online Experiments*” — 2022.
8. Zhou H., Sun K., Li S., Fan Y., Jiang G., Zheng J., Li T. “*STATE: A Robust ATE Estimator of Heavy-Tailed Metrics for Variance Reduction in Online Controlled Experiments.*” — 2024.
9. Google Cloud Public Datasets. “*Google Analytics Sample: Obfuscated Google Analytics 360 Data.*” URL: <https://console.cloud.google.com/marketplace/product/obfuscated-ga360-data/obfuscated-ga360-data>

Дякую за
увагу!