



## РЕПРЕЗЕНТАТИВНІСТЬ ВИБІРОК ПІД ЧАС ВІЙНИ

**ВОЛОДИМИР ПАНІОТТО**

Президент КМІС, професор НаУКМА

Під час війни існують фактори, які погіршують якість даних соціологічних опитувань порівняно з опитуваннями в мирний час. Це масова міграція за кордон і всередині країни (майже третина населення України змінила місце проживання), окупація частини території ворогом, небезпека опитувань у зонах активних бойових дій, зростання рівня патріотизму і, відповідно, вищий тиск переважаючої громадської думки в бік соціально бажаних відповідей, а також прийняття законів, що передбачають кримінальну відповідальність за певні висловлювання.

Головним методом опитувань під час війни в Україні є телефонні опитування за допомогою комп'ютера (CATI). При цьому здійснюється випадкова генерація телефонних номерів та накопичення баз номерів. Частка згенерованих номерів пропорційна до частки мобільних номерів, які припадають на кожного мобільного оператора.

Для оцінки якості даних часто використовують підхід Total Survey Error. Він полягає у виокремленні різних складових загальної помилки та оцінці кожної з них (помилка вибірки, помилка охоплення, тобто помилка, пов'язана з неповним охопленням цільової аудиторії, помилка непридатності відповідей на рівні одиниць тощо). На деякі з цих складових, наприклад, на статистичну помилку вибірки, війна не впливає, проте на інші (наприклад, помилка охоплення або структури вибірки) – впливає значно.

**Два головні питання для оцінки якості соціологічних даних під час війни – чи можна вважати наші дані репрезентативними і чи можемо ми довіряти відповідям тих респондентів, яких опитуємо (проблема ширості відповідей).**

У цій статті я розгляну лише репрезентативність наших даних. Щодо ширості, то два чинники, які найбільше впливають на ширість респондентів – адміністративний тиск (страх переслідування за висловлену думку) і так звана «спіраль мовчання», описана Елізабет-Ноель. Люди не схильні висловлювати свою думку, якщо відчувають, що перебувають у меншості. Методи оцінки і методи підвищення рівня ширості коротко розглянуті у статті: V.Paniotto. Methods for Data Quality Assessment in Wartime Surveys in Ukraine. – [UKRAINIAN ANALYTICAL DIGEST](#) No. 006, May 2024, p. 9-12.

**Отже, репрезентативність опитувань під час війни.** Демографи розрізняють постійне і наявне населення. **Постійне населення** – це ті люди, які зареєстровані й зазвичай проживають на певній території, навіть якщо вони тимчасово відсутні. **Наявне населення** – це люди, які фактично присутні в регіоні на момент опитування, за винятком тимчасово відсутніх постійних жителів. Соціологи зазвичай опитують не постійне, а наявне населення, тобто тих, хто проживає на території України в період опитування (іноді опитуються й аналізуються окремо біженці, але не об'єднуються з наявним населенням). Справа в тому, що для багатьох питань, які вивчають соціологи, опитування біженців не є релевантним (наприклад, доступність Інтернету при відключенні електрики, ставлення до закону про мобілізацію тощо), а більшість питань для біженців (наприклад, чи планують вони повертатися і коли) не є релевантними для наявного населення. Тому нас переважно цікавить репрезентативність опитувань наявного населення, тоді як опитування біженців слід розглядати окремо.

**В Україні проблеми з вибіркою при проведенні соціологічних досліджень існували ще до початку війни, і на це вплинули кілька факторів.**

**По-перше**, відсутність перепису населення протягом 20 років. Останній загальний перепис проводився у 2001 році, і з того часу демографічна структура країни зазнала значних змін. Відсутність актуальних даних ускладнює формування репрезентативної вибірки.

**По-друге**, пандемія COVID-19 внесла значні корективи в методику проведення соціологічних опитувань. До пандемії більшість досліджень проводилася методом особистих інтерв'ю (F2F), але з початком пандемії довелося масово переходити на телефонні опитування (CATI), що має свої обмеження. Зокрема, за останніми достовірними доковідними даними, 4% населення не мали мобільного телефону, а літні люди менш активно користуються мобільними.

**По-третє**, практичне зникнення стаціонарних телефонів (у 2020 році стаціонарні телефони були лише у 7% населення) і перехід від dual frame (комбінація вибірки як мобільних, так і стаціонарних телефонів) до опитувань, заснованих виключно на мобільних телефонах, також ускладнює задачу. Молодь активніше користувалася мобільними телефонами, літні люди – стаціонарними, і ця комбінація давала добрі результати. Однак після переходу на опитування лише за допомогою мобільних телефонів виявилось, що в телефонних опитуваннях на 10% більше людей з вищою освітою і на 10% менше людей з низькими доходами, ніж у F2F інтерв'ю. І зараз неможливо сказати, які дані ближчі до істини – CATI чи F2F інтерв'ю. Попри всі проблеми, у 2019 році основні компанії, що брали участь у виборчих дослідженнях, показали хорошу відповідність опитувань результатам президентських і парламентських виборів (див. наприклад, [Прес-релізи та звіти – Порівняння опитувань та екзит-полів з результатами виборів Президента України 2019 \(1-й тур\) \(kiis.com.ua\)](#)). F2F тоді був точніше, ніж CATI, але за 5 років ситуація могла змінитися.

Для оцінки додаткових систематичних зміщень, які вносять міграція та воєнні дії, необхідно оцінити чисельність тих груп населення, що виникають під час війни і яких важко або неможливо опитати по телефону. **До груп, на які можна поділити населення під час війни, можна віднести такі:**

1. Основна група – населення, яке не змінювало місця проживання після початку війни, ця група доступна для опитування;
2. Комбатанти (учасники бойових дій) – приблизно 2% (хоча вони частково доступні);
3. Населення в зоні активних бойових дій на лінії зіткнення, частково доступні (це 1-2% населення);
4. Біженці (доступні, але важко забезпечити репрезентативне опитування; однак, вони не входять у нашу генеральну сукупність);
5. Внутрішньо переміщені особи (ВПО) – вони повністю доступні для телефонних опитувань;
6. Новоокуповані – населення новоокупованих (після 24 лютого 2022 року) територій: на початку війни їх ще можна було опитувати, але пізніше окупанти відключили українських операторів мобільного зв'язку (це 3-5%);
7. Староокуповані – люди, які проживають на територіях, окупованих раніше, у 2014 році (Крим і частини Луганської та Донецької області); і зараз, і до повномасштабної війни вони не входили в нашу генеральну сукупність.

Таким чином, дані опитувань в Україні репрезентативні для 91-94% населення, яке зараз проживає на території України. Відповідно, **додаткову систематичну помилку репрезентативності в період війни можна оцінити за формулою:**

$$E (\text{error}) = (p1-p2) \times IT \times 0.01$$

де **p1** – процент даної ознаки для населення доступних територій

**p2** – процент даної ознаки для населення недоступних територій

**IT** – (inaccessible territory) процент населення, що проживає на недоступних для опитування територіях

Звісно, ми не знаємо p2, але іноді можемо приблизно оцінити різницю (p1-p2) за довоєнними даними. Наприклад, якщо p1-p2 = 20%, а нам недоступно для опитування 7% населення України, то додаткова помилка репрезентативності E (помилка) = 20% × 7% × 0.01 = 1.4%. Ми бачимо, що навіть якщо різниця відносно велика, 20%, але відсоток недоступного населення невеликий, то додаткова помилка не перевищує 1.5-2%.

Таким чином, помилка вибірки під час війни збільшується, але навіть стандартні методи телефонного опитування зі стохастичною генерацією номерів дозволяють отримати достатньо репрезентативні дані для наявного населення України в цілому. Для отримання даних, репрезентативних для окремих регіонів, необхідно використовувати дані, накопичені у попередніх опитуваннях, репрезентативних для всього населення.

Якщо потрібні більш точні дані (наприклад, для установчого дослідження, яке KMIC проводить для пілпметричної панелі Індустріального телевізійного комітету / медіа груп), використовується комбінація телефонного опитування з використанням методу соціальних мереж для уточнення рівня міграції з F2F інтерв'ю, яке включає перепис точок опитування (складання списку домогосподарств).

Метод оцінки міграції із використанням соціальних мереж я запропонував у 2014 році і опишу його на прикладі оцінки міграції з Донецької та Луганської областей у 2014-2015 роках. Оцінки різних міністерств, зокрема Міністерства надзвичайних ситуацій та Міністерства соціальної політики, щодо кількості внутрішньо переміщених осіб (ВПО) з Донецької та Луганської областей різнилися у 1,5 рази, що вказує на значну неточність офіційної статистики.



Систематична помилка, пов'язана з проблемами репрезентативності та ширості респондентів під час війни, зростає, але не настільки, щоб ставити під сумнів основні результати опитувань.

Однією з проблем у дослідженнях міграції є те, що звичайні соціологічні опитування не дозволяють точно оцінити кількість тих, хто виїхав, бо якщо люди виїхали всією родиною, то вони не включаються до вибірки, це створює викривлення у даних.

Суть нашого підходу полягає у використанні соціальних мереж респондентів для збору інформації про тих, хто виїхав. Ми проводимо стандартну стохастичну вибірку, але запитуємо у респондентів про п'ять сімей, які є для них найближчими та які проживали в Україні до початку війни. Далі респонденти повідомляють, хто з цих сімей виїхав, а хто залишився (див. додаток).

Такий підхід дозволяє отримати інформацію про сім'ї, де всі члени виїхали, а також значно збільшити ефективний розмір вибірки. Наприклад, якщо опитати 2000 респондентів, можна отримати інформацію про приблизно 12,000 осіб, що значно підвищує точність оцінок міграції. Наші оцінки обговорювались експертами замовника, а також на двох конференціях і були визнані адекватними. Ми використали цей метод і в 2023, і в 2024, але виникли певні методичні виклики, які привели до модифікації методу. Актуальні дані (наразі – на 2023 рік) Замовник – ІТК та медіа групи – люб'язно дозволяють використовувати, їх можна отримати в ІТК чи KMICi (а [тут відео](#) презентації результатів).

Загалом можна сказати, що систематична помилка, пов'язана з проблемами репрезентативності та ширості респондентів під час війни, зростає, але не настільки, щоб ставити під сумнів основні результати опитувань. У разі необхідності її можна оцінити в ході методичних експериментів і використовувати методи, які зменшують помилку.

## Додаток. Фрагмент анкети.

### БЛОК ТІЛЬКО ДЛЯ ЖИТЕЛІВ ДОНЕЦЬКОЇ И ЛУГАНСЬКОЇ ОБЛАСТЕЙ

А зараз кілька незвичайних питань. Згадайте, будь ласка, січень 2014 року. Наприклад, як Ви святкували Новий Рік та Різдво у 2014 році (1 січня, 7 січня 2014 року). Тепер згадайте, будь ласка, 5 Ваших найближчих знайомих, які постійно проживають у НАЗВАТИ НАСЕЛЕНИЙ ПУНКТ або жили тут у січні 2014 року. Це мають бути люди із різних сімей. Ми не будемо запитувати Вас, хто вони та ким вони Вам припадають. Тепер по кожному з цих людей маємо 6 питань:

- 1 Ця людина продовжує жити у цьому місті / селі?
- 2 Скільки всього людей було у сім'ї у січні 2014 року, тобто скільки людей, у т.ч. і діти, проживали одним домогосподарством із цією людиною, включаючи цю людину?
- 3 Скільки людей із цієї родини виїхали з населеного пункту та більше не поверталися (зараз тут не проживають)?
- 4 Ці люди поїхали в межах України чи закордон?  
ЯКЩО У МЕЖІ УКРАЇНИ: В яку саме область вони переїхали?
- 5 А скільки людей із цієї родини виїжджали із населеного пункту, але повернулися і зараз знову мешкають тут?
- 6 Ці люди переїжджали в межах України чи закордон?  
ЯКЩО У МЕЖІ УКРАЇНИ: В яку саме область вони переїжджали?



Київський міжнародний інститут соціології – приватна українська дослідницька компанія, заснована у 1992 році. За допомогою мережі інтерв'юерів в усіх областях України проводить особисті інтерв'ю на планшетах. КМІС має можливість залучити одночасно 100-150 інтерв'юерів для проведення телефонних опитувань.

Щорічно реалізує близько 150 дослідницьких проєктів. Основні напрями діяльності: соціо-економічні та політичні дослідження, дослідження в сфері здоров'я, маркетингові дослідження, консалтинг та аудит.



[kiis.com.ua](http://kiis.com.ua)



[office@kiis.com.ua](mailto:office@kiis.com.ua)



+38 044 337 33 76