

РЕПРЕЗЕНТАТИВНІСТЬ МАРКЕТИНГОВИХ ДОСЛІДЖЕНЬ: ПІДХОДИ ТА КРИТЕРІЇ



ЮРІЙ ДАЙНОВСЬКИЙ
 д. е. н., проф., зав. кафедрою маркетингу,
 Львівська комерційна академія
 yddd@ukr.net

У статті проаналізовано підходи до визначення обсягів вибірових сукупностей у маркетингових дослідженнях. Показано недоцільність розрахунку чисельності вибірки виключно на основі широко застосовуваних математичних формул. Запропоновано критерії визначення достатності обсягу вибірки за даними поетапного аналізу нагромадженого емпіричного матеріалу

Забезпечення належної репрезентативності вибірки під час проведення маркетингових досліджень завжди було і залишається одним із найважливіших чинників їхньої ефективності. Зменшення обсягу вибірки переважно знижує витрати на проведення дослідження, проте збільшує вірогідність того, що отримані дані не повною мірою відбиватимуть дійсність і, можливо, спровокують необґрунтовані подальші управлінські рішення. Водночас надмірне збільшення обсягу досліджуваної вибіркової сукупності призводить до необхідності додаткового й інколи значного фінансування, проте не завжди може забезпечити відповідне підвищення точності отримуваних оцінок реального і потенційного попиту, закономірностей розвитку ринку тощо. Крім власне визначення обсягу вибірки, потрібно врахувати низку інших чинників, правильність оцінювання яких урешті відобразиться на репрезентативності отриманих даних. Значна кількість наявних методичних підходів і рекомендацій, математико-статистичний апарат, пропонований для розрахунків обсягу вибірки, не завжди спрощує проблему забезпечення репрезентативності, яка щоразу постає перед дослідниками.

Один із найпопулярніших підходів, що його пропонують для визначення доцільного обсягу вибірки, ґрун-

тується на оцінюванні варіації значень суттєвих для дослідника показників одиниць досліджуваної сукупності та прийнятті попереднього рішення щодо того, яка гранична похибка вибірки могла би бути прийнятною для дослідника [1, 398; 2, 89-91]. Узагальнено сутність цього підходу втілюють відомі формули:

$$n = \frac{t^2 Q^2}{\Delta x^2} \quad (1)$$

$$n = \frac{t^2 pq}{\Delta p^2} \quad (2)$$

де n – необхідний обсяг вибірки;
 x – кількісний вираз величини певної ознаки одиниць досліджуваної сукупності (кількість споживання певного продукту, ціна придбання, термін експлуатації);

Q – середнє квадратичне відхилення конкретних значень кількісної ознаки для одиниць досліджуваної сукупності від їхнього середньоарифметичного значення;

p – частка одиниць досліджуваної сукупності, що характеризуються певною якісною ознакою (частка сімей, які мають автомобілі; частка осіб, які палять; частка помешкань, де встановлено сучасні віконні системи);

q – частка одиниць досліджуваної сукупності, щодо яких зазначеної вище якісної ознаки немає ($q=1-p$);

Δx – максимальна похибка вибірки щодо кількісної ознаки (у натуральних одиницях виміру), що влаштовуватиме дослідника;

Δp – максимальна похибка вибірки щодо якісної ознаки (виражена як частка від одиниці), що влаштовуватиме дослідника. Тобто під час планування дослідження як умова закладається, що відхилення середніх значень вибіркової та генеральної сукупностей не має перевищувати максимальних (граничних) похибок вибірки Δx , Δp ;

t – коефіцієнт довіри, який відбиває рівень гарантованості того факту, що відхилення середніх значень вибіркової та генеральної сукупнос-

Коефіцієнт довіри t	Вірогідність, із якою можна гарантувати, що фактична похибка вибірки не перевищуватиме Δx , Δp
1,0	0,683
1,5	0,866
2,0	0,954
2,5	0,988
3,0	0,997
3,5	0,999

тей справді не перевищить величин Δx і Δr . Теорія вірогідності й математична статистика доводять, що залежність між величиною t і зазначеним рівнем гарантованості є такою:

Виходячи з такої залежності, дослідники пропонують самостійно визначитися щодо того, яку величину коефіцієнта довіри доцільно закласти під час планування дослідження. Водночас він має пам'ятати, що бажання збільшити цей коефіцієнт удвічі вимагає збільшення вибіркової сукупності у 4 рази. Як правило, t приймають рівним 2, виходячи з такої логіки: менші значення не дають достатнього рівня гарантованості, а більші вимагають надмірного збільшення обсягів вибірки.

Зазначений теоретичний підхід створює певний стереотип відносин між дослідником і замовником або опонентами, за якого на запитання "Якою є репрезентативність дослідження?" постає чітка відповідь на кшталт: „Похибка у межах 3%”. Проте аналіз ситуації як з наукових позицій, так і виходячи з практики реальних досліджень дозволяє висловити сумнів у доцільності надання цьому підходові пріоритетного значення у плануванні маркетингових досліджень. Спробуємо обґрунтувати сказане низкою аргументів.

1. На реальні похибки вибірових маркетингових досліджень значно більший вплив, ніж випадкові помилки репрезентативності (на врахування яких власне і спрямовані математичні формули) можуть мати:

- а) помилки реєстрації, пов'язані з неточною фіксацією даних (наприклад, у разі вимірюванні площі помешкання, визначенні "на око" вікової групи, до якої належить респондент);
- б) систематичні помилки, що призводять до зміщення вибірки (наприклад, опитування осіб, які є приємнішими співрозмовниками, більш комунікабельними, й ігнорування тих, із ким важче налагодити контакт). Типовою систематичною помил-

кою вибіркового усного опитування є потрапляння до вибірки менш заклопотаних осіб, які можуть приділити час відповідям на запитання;

- в) брак достеменної основи вибірки (списку представників генеральної сукупності з їхніми координатами), відмінності між характеристиками генеральної сукупності досліджуваних одиниць і сукупності, що приймається за основу вибірки;
- г) недостатня продуманість технології відбирання одиниць спостереження;
- д) недосягнення якісної репрезентативності вибірки внаслідок, наприклад, наявності у ній не всіх груп одиниць спостереження, що існують у генеральній сукупності, або невідповідність структури вибіркової та генеральної сукупності за об'єктивно наявними в них групами осіб чи інших досліджуваних одиниць;
- е) неправильне визначення місця або часу дослідження, що може бути важливим для забезпечення репрезентативності.

2. Коли йдеться про помилку частки, те саме значення похибки вибірки можна оцінювати по-різному залежно від реальної частки певної групи одиниць спостереження у генеральній сукупності. Так, похибка вибірки $\pm 3\%$ є прийнятною, якщо йдеться про репрезентативність частки осіб різної статі серед респондентів (приблизно по 1/2), але вона може принципово вплинути на результат дослідження, якщо стосуватиметься частки мешканців України, які користуються послугами авіаперевізників, адже остання є невеликою і помилка у 3% кардинально змінить її величину. Оскільки значна кількість запитань анкети передбачає певні переліки альтернатив відповідей, серед них обов'язково будуть такі, що матимуть різну питому вагу у відповідях респондентів. Отож, той самий відсоток похибки щодо різних альтернатив

відповідей свідчитиме про різний рівень репрезентативності.

3. Переважно дослідження передбачає отримання відповідей не на одне, а на низку запитань. У такому випадку щодо кожного питання необхідно проводити окремий розрахунок кількості вибірки, а при конкретній фактичній кількості досліджених одиниць вибіркової сукупності помилка реально буде різною щодо різних питань або досліджуваних показників.

4. Величини Q і q мали би бути відомими до проведення дослідження, адже вони потрібні для розрахунку кількості вибірки. Однак реально ці величини можна виявити лише за результатами принаймні окремих етапів дослідження. Крім того, ці величини у будь-якому разі стосуватимуться дослідженої вибіркової сукупності, а не генеральної сукупності одиниць дослідження, що теоретично передбачено у разі застосування формул (1) і (2).

5. Якщо вибірка загалом є репрезентативною, то це не означає, що репрезентативними будуть і дані групувань. Під час групування відбувається подрібнення матеріалу, й похибки вибірки в окремих групах виявляються вищими, ніж у вибіркової сукупності загалом.

6. Використання наведених формул ґрунтується на припущенні про нормальний розподіл величин досліджуваних ознак одиниць генеральної сукупності. За нормального розподілу правий і лівий боки кривої розподілу є симетричними, середнє значення ознаки збігається з модою і медіаною. 68,3% одиниць спостереження матимуть значення досліджуваної ознаки x у межах $\pm Q$, 95,4% – у межах $\pm 2Q$, 99,7% – у межах $\pm 3Q$. Для переважної більшості кривих розподілу реально досліджуваних показників такі умови не дотримуються, а це означає, що розрахунки за формулами (1) і (2) втрачають сенс. Наприклад, значна кількість досліджуваних показників може мати тільки додатне значення

(доходи респондентів, кількість наявних у помешканні товарів певного виду, середньодобовий кілометраж пробігу автомобілів). Це означає, що ліва частина кривої розподілу обмежується нульовим значенням, що не може не впливати на конфігурацію правої частини. А отже, загалом фактична крива розподілу відрізняється від теоретичної кривої нормального розподілу, що знову-таки не дозволяє використовувати формули (1) і (2).

У всіх фундаментальних працях із загальної теорії статистики, теорії вірогідності та математичної статистики підкреслюється, що необхідною умовою нормального розподілу є відсутність домінуючих чинників, що впливають на відхилення від середньої [3, 201]. Крім того, розподіл ознаки може бути нормальним, якщо ця ознака є наслідком впливу численних і різноманітних чинників, які мало пов'язані один з одним і вплив кожного з яких є малим порівняно із загальним впливом усіх чинників [3, 210]. У переважній більшості соціально-економічних процесів і показників є чинники, вплив яких якщо не домінує, то принаймні є вельми суттєвим, причому чинники взаємопов'язані, тобто нормального розподілу навіть теоретично не повинно бути. Однак про це чомусь забувають, коли виникає можливість застосувати формули, виведені саме на підставі закономірностей нормального розподілу. Теорія доводить, що навіть сам розрахунок середньоквадратичного відхилення для несиметричних розподілів не має сенсу [3, 192], але й це у практичних розрахунках під час проведення маркетингових досліджень часто ігнорують.

Особливістю розподілу значень багатьох показників, на виявлення яких спрямовані маркетингові дослідження, є те, що "хвости" кривої їхнього розподілу можуть суттєво виходити за межі $\pm 3\sigma$. Адже крайні значення показників часто пов'язані із "захопленням" до вибіркової сукупності, крім основної маси одиниць

досліджуваної сукупності, ще й певної кількості одиниць із суміжних сукупностей, що якісно відрізняються від цільової генеральної. Наприклад, за вибіркового дослідження середніх цін за квадратний метр житлової площі це будуть як помешкання типу маєтки-люкс, так і частини гуртожитків, площ без вигод, переведені у склад житлового фонду. У таких ситуаціях "хвости" емпіричних кривих чи полігонів розподілу можуть не тільки не наближатися до нуля, а й створювати додаткові "горби", засвідчуючи існування певних особливих груп

одиниць спостереження.

Наведені міркування свідчать про наявність значної кількості чинників, що загалом створюють вагоміший вплив на репрезентативність вибіркової сукупності, ніж вплив тих закономірностей, які визначаються суто теоретичними аспектами нормального розподілу і випадкового відбору. Це ставить під сумнів практичну корисність попередніх розрахунків обсягів вибірки на підставі формул (1), (2) та окремих їхніх модифікацій. Не випадково спеціалісти зазначають, що "формули відбивають тільки ви-

Таблиця 1

Приклад залежності варіації середніх субвибіркових значень досліджуваних показників від обсягу вибіркової сукупності

Обсяг вибірки, одиниць	Обсяги субвибірок, одиниць (25% від обсягу вибірки)	Розмах варіації середніх значень показників у субвибірках, % до середнього значення показника у вибірці	Коефіцієнт варіації, %
40	10	19,51	7,33
80	20	11,93	4,23
160	40	4,63	1,66
320	80	2,87	1,14
640	160	2,61	1,06
1280	320	2,12	0,86
1541	385, 386	1,79	0,67

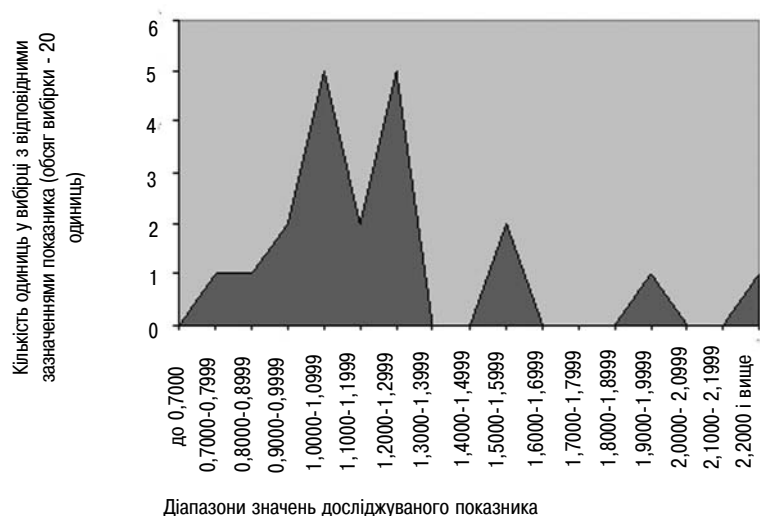


Рис. 1. Полігон розподілу значень досліджуваного показника за матеріалами обстеження 20 одиниць вибіркової сукупності



Рис. 2. Полігон розподілу значень досліджуваного показника за матеріалами обстеження 154 одиниць вибіркової сукупності.

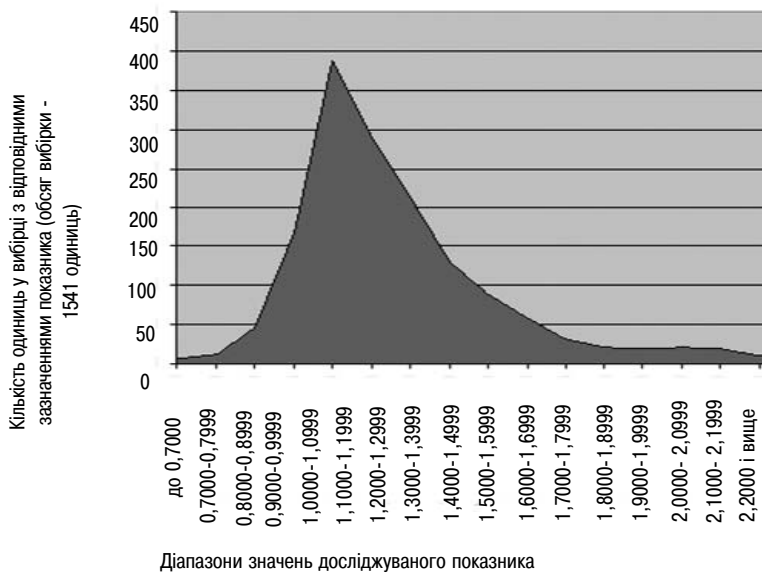


Рис. 3. Полігон розподілу значень досліджуваного показника за матеріалами обстеження 1541 одиниці вибіркової сукупності.

падкові помилки, що, як показує досвід, бувають найменшими з усіх помилок, які трапляються” [4, 16]. Цим можна пояснити й акцентування в окремих наукових працях не методів визначення кількості вибірки, а питань логіки і технології проведення вибіркового дослідження [4] та рекомендацій щодо спрощеного визначення обсягу вибірки на підставі попередньої оцінки загальної кількості мікрогруп, що мають бути

представлені у вибіркової сукупності, з урахуванням необхідності забезпечити певний мінімальний рівень репрезентативності у кожній із цих мікрогруп [1, 399].

На наш погляд, для багатьох досліджень доцільнішим підходом є проведення оцінки достатності або недостатності обсягу вибірки у міру поетапного нагромадження емпіричного матеріалу. Основними критеріями достатності вибірки можуть

бути:

1) варіація середніх значень досліджуваних ознак у субвибіркових сукупностях, сформованих на підставі випадкового відбору в межах фактично охопленої вибірки (якщо варіація є незначною, то субвибірки подібні та їхнім середнім характеристикам можна довіряти; при цьому об’єднання достатньо однорідних субвибірок у загальну вибірку створює додаткові гарантії репрезентативності);

2) плавність конфігурації фактичної кривої розподілу значень досліджуваних ознак у сформованій вибірковій сукупності (репрезентативній вибірці у переважній більшості випадків повинна відповідати плавна крива розподілу).

Проілюструємо такий підхід на прикладі аналізу цінових пропозицій на ринку житла, проведеного у м. Львові у 2007 р. в межах дослідження маркетингової діяльності посередницьких підприємств на цьому ринку [5]. Опрацювання 1541 пропозиції щодо продажу житла дозволила отримати настільки репрезентативну вибірку, що її аналіз виявився, на наш погляд, корисним не тільки щодо реального предмету дослідження, а й із погляду можливості проілюструвати на цьому прикладі процес впливу збільшення вибірки на підвищення її репрезентативності. Тому в наведеному нижче матеріалі ми оперуватимемо неіменований *досліджуваним показником*, оскільки реальна назва й величини цього показника — середньої вартості пропонованого для продажу житла у розрахунок за 1 кв. м загальної площі приміщень в умовних одиницях (доларах США) у Львові у 1 кварталі 2007 р. — не мають значення з точки зору мети даної статті. Із зазначеної вибірки методом випадкового відбору формуватимемо субвибірки різних обсягів, починаючи від малих із послідовним їх збільшенням.

На першому кроці за допомогою випадкового відбору сформовано вибірку обсягом у 40 одиниць, яку розподілено так само за допомогою

випадкового відбору на 4 субвибірки по 10 одиниць. Конкретний спосіб відбору не має значення, але він повинен забезпечити умову випадковості відбору, тобто рівної апріорної вірогідності потрапляння до вибірки для кожної одиниці сукупності. Варто додати, що під час поділу 40 одиниць на 4 групи по 10 одиниць може бути прийнятним і механічний відбір (перша субвибірка – одиниці №№ 1, 5, 9 тощо з вибіркової сукупності; друга субвибірка – одиниці №№ 2, 6, 10 тощо), але тільки у тому разі, якщо вихідна послідовність цих 40 одиниць відповідає вимогам випадковості. Наприклад, такий розподіл на субвибірки буде прийнятним, якщо вибірка із 40 респондентів розміщена в алфавітному порядку їхніх прізвищ. Проте такий відбір не буде коректним, якщо одиниці вибірки проранговано за певним кількісним показником, адже це сприятиме штучному вирівнюванню середніх значень показників у субвибірках. У нашому прикладі середні значення досліджуваного показника у субвибірках становили: 1,4415; 1,1890; 1,3096; 1,2395. Середнє значення для вибірки із 40 одиниць – 1,2949. Розмах варіації середніх субвибіркових значень відносно середнього значення у вибірці становить $(1,4415 - 1,1890) : 1,2949 \times 100\% = 19,5\%$; коефіцієнт варіації, розрахований як процентне відношення середньоквадратичного відхилення до середнього значення, – 7,33% (табл. 1). Якщо дослідник бажає зменшити цю варіацію й таким чином підвищити достеменність отримуваних даних, він може послідовно збільшувати кількість вибірки. У нашому прикладі збільшення досліджуваної сукупності до 80, 160, 320 тощо одиниць призводило до систематичного зменшення показників варіації.

Кількість кроків такого поетапного збільшення вибірки дослідник повинен визначити експертно, виходячи з цілей дослідження. Якщо середнє значення показника коливається у прийнятних межах, ре-

презентативність можна сприймати як достатню. У наведеному прикладі високу репрезентативність вибірки щодо середньої величини досліджуваного показника можна констатувати вже за наявності 160 обстежених одиниць. Проте якщо дослідника цікавить і репрезентативність розподілу значень показника за одиницями сукупності, доцільніше застосувати наступний із запропонованих критеріїв і викреслити полігони розподілу значень показника для різних обсягів вибірки.

На підставі даних дослідження побудовано полігони розподілу значень досліджуваного показника при різній кількості одиниць у вибірковій сукупності: 20 (мала вибірка), 154 (десята частина від усіх досліджених одиниць) і 1541 (весь обсяг зібраної інформації). Вибірка у 20 одиниць лише орієнтовно показує загальний можливий діапазон розсіяння значень досліджуваного показника і діапазон значень, які, вірогідно, траплятимуться дещо частіше; говорити про репрезентативність такої вибірки є явно передчасним (рис. 1). Вибірка у 154 одиниці показує загальну конфігурацію розподілу, хоча наявність кількох піків не дозволяє довіряти даним щодо окремих діапазонів значень (рис. 2). Нарешті полігон, отриманий на підставі даних усіх обстежених одиниць (рис. 3), демонструє абсолютно чітку картину розподілу і дає достатні підстави для висновку про репрезентативність як вибірки загалом, так і про можливість довіряти даним певних необхідних групувань залежно від величини досліджуваного показника.

Таким чином, на наш погляд, раціональний підхід до забезпечення репрезентативності вибірки під час проведення маркетингових досліджень має бути спрямований на таке.

По-перше, це зосередження уваги на логіці рішень щодо забезпечення якісної репрезентативності вибірки з урахуванням способів відбору, місця і часу дослідження, усіх груп, типів,

категорій об'єктів, що у вибірковій сукупності мають (або не мають) бути представленими. Жодні зусилля щодо забезпечення кількісної репрезентативності вибірки не можуть компенсувати помилки у рішеннях, спрямованих на забезпечення якісної репрезентативності.

По-друге, під час визначення доцільної кількості досліджуваних одиниць варто більше орієнтуватися не на формальні математико-статистичні критерії, а на поетапний аналіз отримуваної інформації. При цьому пропонуємо застосовувати критерії досягнення подібності середніх характеристик випадковим чином виокремлених частин фактично сформованої вибірки, а також плавності полігонів розподілу значень показників досліджуваних одиниць. На підставі відповідності отримуваних даних цим критеріям можна приймати експертні рішення щодо доцільності або недоцільності подальшого збільшення обсягів вибірки.

Джерела

1. Скибінський С. В. Маркетинг. Ч. 1 : підручник / Скибінський С. В. – Львів : Місіонер, 2000. – 640 с.
2. Федорчук А. І. Маркетингові дослідження : навч. посібник / Федорчук А. І. – Львів : Видавництво Львівської комерційної академії, 2008. – 368 с.
3. Кильдишев Г. С. Общая теория статистики : учебник / [Кильдишев Г. С., Овсиенко В. Е., Рабинович П. М., Рябушкин Т. В.]. – М. : Статистика, 1980. – 423 с.
4. Чурилов М. Структура процесу формування вибіркової сукупності / Чурилов М. // Маркетинг в Україні. – 2007. – № 4. – С. 14-21.
5. Дайновський Ю. А. Маркетингові складники конкурентоспроможності послуг посередницьких підприємств на ринку житла / Дайновський Ю. А. // Маркетинг в Україні. – 2007. – № 4. – С. 54-60.