
МЕТОДИКА

розрахунку системи статистичних ваг для поширення результатів вибіркового обстеження населення (домогосподарств) з питань економічної активності на генеральну сукупність

1. Основні принципи поширення результатів вибірових обстежень населення (домогосподарств) з питань економічної активності на генеральну сукупність

Державні статистичні спостереження – вибіркові обстеження населення (домогосподарств) з питань економічної активності в Україні проводяться на основі опитування сукупності неінституційних домогосподарств, що формується за процедурою стратифікованого багатоступеневого випадкового відбору. З 2004 року обсяг щомісячної вибірки становить 18,5 тис. домогосподарств, з яких 7,4 тис. домогосподарств відбираються з вибіркової сукупності домогосподарств обстеження сільськогосподарської діяльності в сільській місцевості. При цьому вибірка сукупність домогосподарств представлена шістьма приблизно рівними частинами – ротаційними групами, кожна з яких є репрезентативною вибіркою з генеральної сукупності. Щомісяця дві ротаційні групи (третина вибіркової сукупності) за спеціальною схемою заміщуються двома іншими групами – однією новою та ротаційною групою, що була обстежена у відповідному місяці минулого року [3].

У кожному домогосподарстві, яке взяло участь в обстеженні, збирається демографічна інформація по всіх членах домогосподарства і спеціальна інформація щодо економічної активності, зайнятості та безробіття членів домогосподарств віком 15-70 років.

Для науково обґрунтованого поширення результатів ОЕАН на генеральну сукупність розраховується система статистичних ваг обстежених осіб. Статистичні ваги призначені для врахування загальних імовірностей включення домогосподарств до вибірки, врахування фактичних рівнів відмов домогосподарств та окремих осіб від участі в обстеженні, гармонізації результатів обстеження з даними демографічної статистики щодо чисельності та статеві-вікової структури населення тощо [4-6].

У загальному випадку кінцева вага i -ої особи, обстеженої в ОЕАН, розраховується як добуток базової ваги домогосподарства та відповідних вагових коефіцієнтів (коефіцієнтів коригування ваг) за формулою:

$$w_i = w_{B_i} \cdot k_{1i} \cdot \dots \cdot k_{Ei}, \quad (1)$$

де:

w_i – кінцева вага особи;

w_{B_i} – базова вага домогосподарства, членом якого є особа;

$k_{1i}, \dots, k_{Ei}$ – вагові коефіцієнти;

E – кількість етапів коригування статистичних ваг.

Система вагових коефіцієнтів повинна відображати ефекти впливу всіх етапів коригування ваг на кінцеві ваги осіб.

Для ОЕАН у 2004-2008 роках розрахунок системи статистичних ваг складається з наступних основних етапів:

- розрахунок базових ваг домогосподарств;
- коригування базових ваг для врахування відмінностей між теоретичною та фактичною кількостями відібраних домогосподарств;
- коригування ваг для врахування ступеня участі домогосподарств в обстеженні;
- коригування ваг для врахування ступеня участі осіб в обстеженні;
- коригування ваг для узгодження результатів обстеження з наявною зовнішньою інформацією;
- контроль якості системи ваг.

Етап розрахунку системи базових ваг домогосподарств є визначальним. Як уже відмічалось, призначення базових ваг – це врахування реального дизайну вибірки – принципів побудови, механізмів відбору тощо.

Базова вага є результатом певної послідовності розрахунків. Дизайн вибірки для ОЕАН у 2004 – 2008 роках обумовлює використання для побудови базових ваг наступних складових [3]:

- імовірність відбору одиниць першого ступеню P_{1i}^1 ;
- імовірність відбору одиниць другого ступеню P_{2i}^2 ;
- імовірність відбору домогосподарств у межах одиниць вибірки другого ступеню по міських поселеннях P_{3i} .

Базовою вагою домогосподарства, до якого належить i -та особа, є обернена величина добутку зазначених імовірностей:

$$w_{Bi} = \frac{1}{p_i} = \begin{cases} \frac{1}{P_{1i} \cdot P_{2i} \cdot P_{3i}}, & \text{по міських поселеннях} \\ \frac{1}{P_{1i} \cdot P_{2i}}, & \text{по сільській місцевості} \end{cases}, \quad (2)$$

В ОЕАН величина загальної імовірності відбору p_i визначається виходячи з того, що кожне домогосподарство при відборі повинне мати однакову ймовірність потрапити до вибірки. Вибірка, яка побудована з урахуванням зазначеної умови, є самозваженою. Для забезпечення самозваженості вибірки ОЕАН величина p_i визначається за наступною формулою:

$$p_i = f = \begin{cases} P_{1i} \cdot P_{2i} \cdot P_{3i}, & \text{по міських поселеннях} \\ P_{1i} \cdot P_{2i}, & \text{по сільській місцевості} \end{cases}, \quad (3)$$

де f – частка відбору домогосподарств, яка для самозваженої вибірки дорівнює відношенню її обсягу до загальної кількості домогосподарств у генеральній сукупності ($f = \frac{n}{N}$). Величина, обернена до f , дорівнює кількості домогосподарств генеральної сукупності, яку репрезентує одне відібране домогосподарство при плануванні вибірки.

У відповідності з формулою (3) кількість домогосподарств, які відбираються на останньому етапі формування вибірки по територіальній одиниці h , до якої належить i -та особа, $(n_{h(i)})$ визначається за наступним теоретичним співвідношенням:

¹ Одиницями відбору першого ступеню є міські поселення – окремі міста (міськради) або селища міського типу (селищні ради) та сільські ради по сільській місцевості.

² По міських поселеннях це інструкторські дільниці (або сукупності інструкторських дільниць), сформовані для проведення Всеукраїнського перепису населення 2001 року, у сільській місцевості – домогосподарства.

$$n_{h(i)} = \begin{cases} \frac{p_i \cdot N_{3i}}{P_{1i} \cdot P_{2i}} = \frac{f \cdot N_{2i}}{P_{1i} \cdot P_{2i}}, & \text{по міських поселеннях} \\ \frac{p_i \cdot N_{2i}}{P_{1i}} = \frac{f \cdot N_{1i}}{P_{1i}}, & \text{по сільській місцевості} \end{cases}, \quad (4)$$

де:

N_{2i} – загальна кількість домогосподарств у територіальній одиниці другого ступеню, до якої належить i -те домогосподарство, по міських поселеннях;

N_{1i} – загальна кількість домогосподарств у територіальній одиниці першого ступеню, до якої належить i -те домогосподарство, по сільській місцевості.

Базові ваги розраховуються на весь термін експлуатації вибірки. Територіальна вибірка для ОЕАН розрахована на п'ять років: 2004–2008 роки.

Теоретична кількість одиниць, яку необхідно відібрати по кожній територіальній одиниці при побудові вибірки і яка забезпечує самозваженість вибірки, залишається саме теоретичною, оскільки на практиці доводиться враховувати такі фактори, як умови роботи статистиків з проведення вибірових обстежень населення, домогосподарств (далі – статистиків з обстежень)¹, очікувані рівні участі домогосподарств та осіб в ОЕАН, а також виправляти певні похибки в основі вибірки або у процедурах відбору. Тому необхідно скоригувати базові ваги з урахуванням фактичної кількості відібраних домогосподарств. Це здійснюється шляхом введення спеціального коефіцієнту коригування ваг для врахування фактичної кількості відібраних домогосподарств.

Якість результатів ОЕАН може суттєво погіршуватися внаслідок впливу такого явища, як відмови домогосподарств та осіб від обстеження. Це явище призводить до викривлення параметрів вибіркової сукупності, що побудована як самозважена, у порівнянні з параметрами генеральної сукупності. Зокрема, при обстеженнях населення у вибірці спостерігається надлишок представників старших вікових груп обох статей і не вистачає представників молодших вікових груп, серед яких кількість відмов значно вища і, крім того, яких важче застати вдома. Якщо не враховувати ці викривлення, то, відповідно, всі характеристики домогосподарств будуть зміщені в бік домогосподарств, що включають осіб старшого та похилого віку або складаються лише з таких осіб. Особливо це стосується територій зі значними рівнями відмов.

Одним з найбільш розповсюджених і теоретично обґрунтованих методів врахування випадків недержання відповідей при обробці даних вибірових обстежень є коригування статистичних ваг, сутність якого полягає в завищенні ваг одиниць, які мають характеристики, аналогічні характеристикам одиниць, що не були обстежені.

Коригування ваг для компенсації відмов включає розділення вибірки на систему комірок (класів зважування) та розрахунок коефіцієнту для коригування ваг одиниць по кожній комірці як частки відповідей. Процедура коригування ваг включає наступні кроки:

- розділення вибіркової сукупності домогосподарств на групи (комірки зважування), що охоплюють всю цю сукупність і не перетинаються;
- обчислення ступеня участі одиниць в обстеженні для кожної комірки;
- визначення єдиного коефіцієнту коригування ваг для всіх одиниць комірки як величини, оберненої до рівня участі домогосподарств в обстеженні по комірці.

Комірки для коригування ваг формуються як за змінними дизайну вибірки (тобто на основі зовнішньої інформації, що використовується при формуванні

¹ Наприклад, забезпечувати їх рівномірну завантаженість.

вибірки, прикладами якої є: адміністративно-територіальний поділ України, тип населених пунктів тощо), так і за змінними, визначеними за результатами обстеження (тип домогосподарства, кількість членів домогосподарства тощо).

Для узгодження результатів вибіркового обстеження з наявною додатковою інформацією щодо певних характеристик генеральної сукупності здійснюється спеціальне коригування системи статистичних ваг.

Доцільність узгодження результатів обстеження з зовнішньою інформацією обумовлюється такими обставинами:

- 1) оцінки показників за результатами вибіркового обстеження завжди характеризуються певною похибкою, джерелами якої є несуттєвість обстеження (похибка вибірки) і наявність так званих невібіркових похибок;
- 2) вибіркове обстеження населення не може забезпечити оцінок певних характеристик генеральної сукупності, наприклад, оцінки загальної чисельності населення або загальної кількості домогосподарств. Хоча ці оцінки й отримуються за результатами обстеження, вони відображають головним чином параметри, що були закладені при формуванні вибірки;
- 3) основні характеристики генеральної сукупності змінюються у часі і відповідно окремі з них на момент завершення обробки даних обстеження дещо відрізняються від таких, якими вони були під час організації обстеження.

Найбільш обґрунтованим шляхом вирішення всіх цих проблем є коригування статистичних ваг домогосподарств. Статистико-математичні методи, що застосовуються в теперішній час для коригування системи статистичних ваг з метою узгодження результатів обстежень з додатковою інформацією, мають загальну назву методів калібрації [7]. Калібрація системи статистичних ваг полягає у її трансформації (коригуванні) шляхом вирішення математичної задачі мінімізації спеціальної функції відстані між калібраційними вагами та вагами, що коригуються, за умови, що калібраційні ваги забезпечують оцінки показників для генеральної сукупності або її частин, які відомі з зовнішніх даних. Окремим спрощеним варіантом калібрації є метод постстратифікації, який застосовується у випадку, коли наявні надійні зовнішні дані по певних стратах (комірках) генеральної сукупності, які охоплюють всю сукупність і не перетинаються. Зазначені страти, як правило, відрізняються від страт, що були побудовані при формуванні вибіркової сукупності. За таких умов коефіцієнти для коригування системи ваг визначаються по кожній страті окремо, як відношення значення показника за зовнішніми даними до оцінки показника за результатами обстеження або як коефіцієнти, що приводять структуру генеральної сукупності за результатами обстеження до її структури за зовнішніми даними.

За останні роки розроблено методологічні основи, методики, процедури, що дозволяють ефективно коригувати статистичні ваги для забезпечення узгодженості результатів обстеження з додатковою інформацією, накопичено значний досвід з питань аналізу зовнішніх джерел даних.

Для постстратифікації статистичних ваг в обстеженнях економічної активності використовуються, насамперед, дані демографічної статистики щодо чисельності, розміщення та статеві-вікової структури населення у працездатному віці.

Важливою перевагою використання системи ваг є те, що це дозволяє об'єднувати результати різних вибіркових обстежень на мікрорівні. При обстеженні населення з питань економічної активності об'єднуються щомісячні дані, отримані за різними вибірками. А саме дані, отримані на основі вибірки, побудованої виключно для цілей ОЕАН, з даними, отриманими по сільських домогосподарствах, що крім ОЕАН підлягають обстеженню за програмою вибіркового обстеження їх сільськогосподарської діяльності (ОСГД). Це вимагає відповідного коригування системи ваг.

Для ідеальної самозваженої вибірки (при ідеальній реалізації всіх процедур відбору та всіх етапів обстеження) потреби в аналізі системи ваг немає, оскільки ваги всіх одиниць вибірки однакові. На практиці вибірка завжди відхиляється від стану самозваженості за рахунок таких факторів, як наявність у списках для відбору адрес неіснуючих домогосподарств, відхилення навантаження статистиків з обстежень від теоретично визначеної величини, відмов домогосподарств та одиниць від обстеження, що мають різний рівень в різних територіальних одиницях, тощо. У зв'язку з впливом ваг на якість оцінок показників постає нестандартна проблема жорсткого контролю за якістю самих ваг. У державних обстеженнях населення необхідно аналізувати якість системи ваг і використовувати всі можливості для її поліпшення.

У відповідності з наведеними принципами, побудовано порядок розрахунку системи ваг ОЕАН, який схематично представлений на рис. 1.

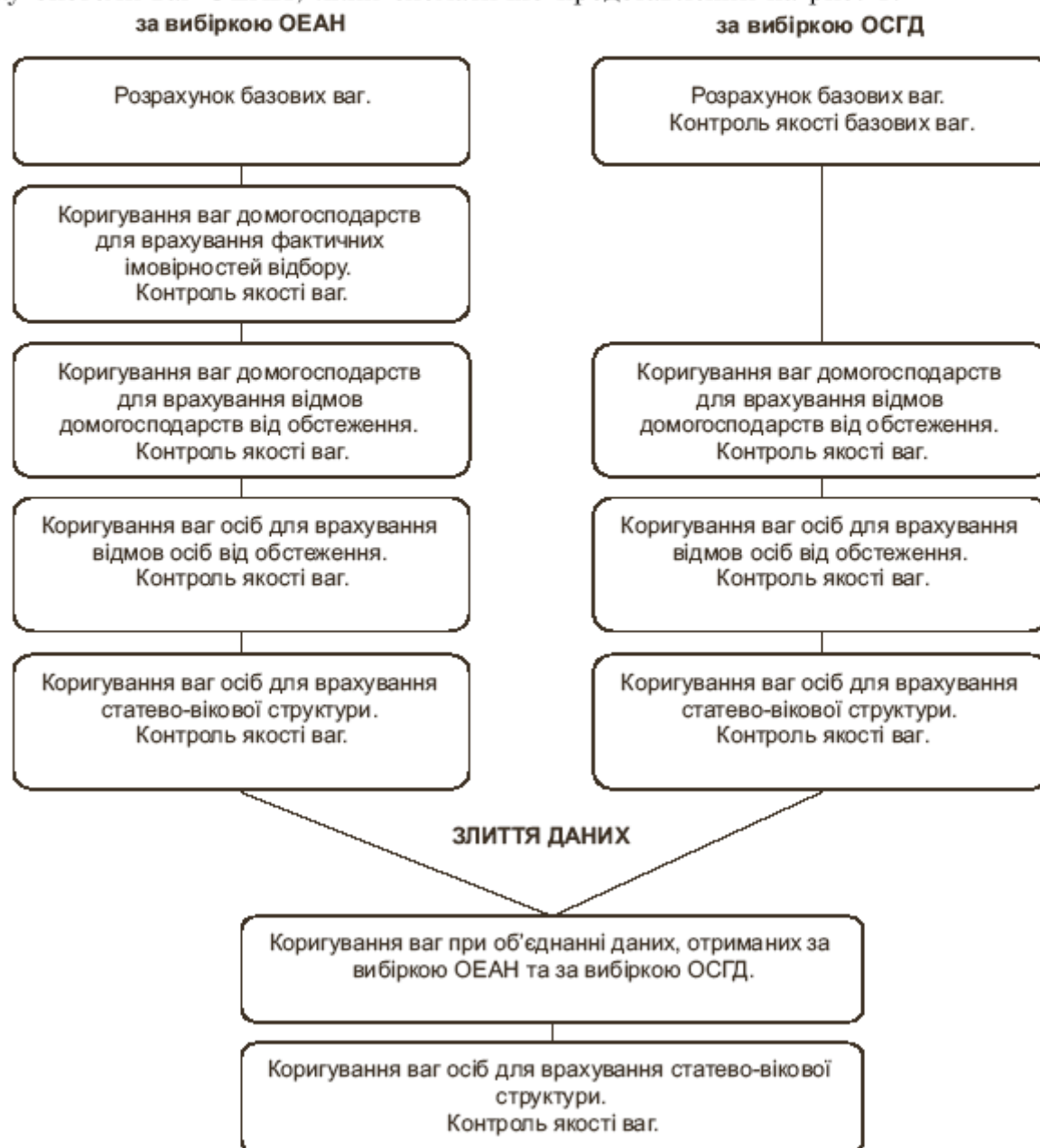


Рис. 1. Етапи розрахунку статистичних ваг для ОЕАН

2. Порядок розрахунку складових системи ваг ОЕАН

2.1. Розрахунок базової ваги домогосподарств

2.1.1. Розрахунок базової ваги домогосподарств міських поселень Визначення імовірності відбору одиниць першого ступеня $P_{\text{и}}$

А. Саморепрезентативні міські поселення

Великі міста

Саморепрезентативними є одиниці відбору першого ступеня з розміром (чисельністю населення) вище встановленого порогу саморепрезентативності. Для таких одиниць відбору $P_{\text{и}} = 1$. Порогове значення чисельності n_p визначається за формулою:

$$n_p = \frac{N}{n} \cdot m \cdot m_f, \quad (5)$$

де:

n – обсяг квартальної вибірки, домогосподарств;

m – квартальне навантаження статистика з обстежень (кількість домогосподарств, яка має бути обстежена одним статистиком з обстежень протягом одного кварталу), домогосподарств;

m_f – середній розмір домогосподарства, осіб;

N – обсяг генеральної сукупності, домогосподарств.

Для міських поселень з чисельністю населення 100 тис. осіб параметри ОЕАН мають наступні значення: загальна кількість неінституційних домогосподарств в Україні $N = 17609000$; обсяг вибірки ОЕАН $n_{\text{ОЕАН}} = 31800$; навантаження статистика з обстежень в ОЕАН $m_{\text{ОЕАН}} = 72$ домогосподарства; середній розмір домогосподарства $m_f = 2,62$ особи. Поріг саморепрезентативності:

$$n_p = 553,75 \cdot 72 \cdot 2,62 \approx 104500 \text{ осіб} \quad (6)$$

Таким чином, міські поселення з чисельністю населення $n \geq 104,5$ тис. чол. (саморепрезентативні міста) включалися у вибірку з імовірністю, рівною одиниці.

Малі міста

Враховуючи, що за принципами організації ОЕАН квартальне навантаження статистиків з обстежень у малих містах нижче, ніж у великих, відповідно 54 і 72 домогосподарства, то необхідно виділити саморепрезентативні малі міста, які також включаються до вибірки з імовірністю рівною одиниці. Для малих міст $m_{\text{ОЕАН}} = 54$ домогосподарства; середній розмір домогосподарства $m_f = 2,65$ особи.

Поріг саморепрезентативності n_{p2} :

$$n_{p2} = 553,75 \cdot 54 \cdot 2,65 \approx 79242 \text{ осіб} \quad (7)$$

Таким чином, у вибірці 2004-2008 років для ОЕАН представлені всі міські поселення з чисельністю населення більше 79 тисяч осіб.

Б. Несаморепрезентативні міські поселення

Несаморепрезентативні міські поселення відбирались по кожному регіону (по кожній страті) окремо. Кількість несаморепрезентативних міст (a), яку необхідно відібрати у регіоні, визначалась шляхом ділення чисельності населення всіх несаморепрезентативних міст регіону на навантаження статистика з обсте-

жень в цих містах ($m_{\text{ОЕАН}} = 54$ домогосподарства) та середній розмір домогосподарств в них ($m_f = 2,65$ особи).

Для міських поселень з населенням менше 79 тисяч осіб імовірність відбору у регіоні розраховується за формулою:

$$P_{1i} = \left(a \cdot \frac{M_\alpha}{\sum_\alpha M_\alpha} \right), \quad (8)$$

де:

a – кількість відібраних малих міст у регіоні;

M_α – чисельність населення міста, для якого розраховується P_{1i} ;

$\sum_\alpha M_\alpha$ – загальна чисельність населення несамопрезентативних малих міст у регіоні.

Визначення імовірності відбору одиниць другого ступеня P_{2i}

Імовірність відбору одиниць другого ступеня P_{2i} (по міських поселеннях – це території інструкторських ділянок, створених для проведення Всеукраїнського перепису населення 2001 року) розраховується за формулою:

$$P_{2i} = \left(b \cdot \frac{M_\beta}{\sum_\beta M_\beta} \right), \quad (9)$$

де:

b – кількість ВТОВ, відібраних у конкретному місті;

M_β – чисельність постійного населення за даними перепису у ВТОВ, для якої розраховується P_{2i} ;

$\sum_\beta M_\beta$ – загальна чисельність населення у відбраному місті.

Якщо розміри територіальних одиниць другого ступеня суттєво відрізняються у межах міста, може виникнути ситуація, коли ймовірність відбору P_{2i} буде дорівнювати або перевищувати одиницю. У такому випадку ймовірність відбору ВТОВ приймається за одиницю, її чисельність населення віднімається від загальної чисельності населення міста, а при розрахунку ймовірності відбору інших ВТОВ, b зменшується на одиницю. Тобто, така велика за чисельністю населення ВТОВ враховується як саморепрезентативна.

Загальна імовірність включення до вибірки ВТОВ розраховується як добуток імовірностей P_{1i} і P_{2i} для кожної ВТОВ. При формуванні вибірки необхідно аналізувати загальні імовірності відбору ВТОВ для виявлення можливих похибок, які відображаються, зокрема, у вигляді окремих нестандартних значень (аутлайерів), що суттєво відрізняються від середнього значення імовірності і можуть призвести до погіршення якості системи ваг.

У випадку виявлення аутлайерів застосовуються спеціальні процедури їх коригування, зокрема, процедури імпутації.

Визначення імовірності відбору домогосподарств P_{3i}

Теоретична імовірність відбору i -го домогосподарства в межах відбраної ВТОВ розраховується, виходячи з формули:

$$P_{3i} = \frac{f}{(P_{1i} \cdot P_{2i})}, \quad (10)$$

де f – частка відбору домогосподарств

За таких умов базова вага, однакова для всіх домогосподарств, визначається за формулою:

$$w_{Bi} = \frac{1}{P_{1i} \cdot P_{2i} \cdot P_{3i}} = \frac{1}{f} \quad (11)$$

2.1.2. Розрахунок базової ваги домогосподарств сільської місцевості Визначення імовірності відбору одиниць першого ступеня P_{1i}

Імовірність включення до вибірки одиниць першого ступеня відбору у сільській місцевості (сільських рад) P_{1i} по кожному району розраховується за формулою:

$$P_{1i} = \left(c \cdot \frac{M_c}{\sum_c M_c} \right), \quad (12)$$

де:

c – кількість відібраних сільських рад у районі;

M_c – кількість домогосподарств відібраної сільської ради, для якої розраховується P_{1i} ;

$\sum_c M_c$ – загальна кількість домогосподарств у сільському районі.

Визначення імовірності відбору домогосподарств P_{2i}

Теоретична імовірність відбору i -го домогосподарства відібраної сільської ради розраховується за формулою (див. формулу 3):

$$P_{2i} = \frac{f}{P_{1i}} \quad (13)$$

За таких умов базова вага, однакова для всіх домогосподарств, дорівнює:

$$w_{Bi} = \frac{1}{P_{1i} \cdot P_{2i}} = \frac{1}{f} \quad (14)$$

2.1.3. Розрахунок частки відбору домогосподарств f

При формуванні вибіркової сукупності домогосподарств для ОЕАН у 2004 році частка відбору дорівнювала $f = \frac{n_{\text{ОЕАН}}}{N} = \frac{31800}{17609200}$. Ця величина розрахована як відношення обсягу вибіркової сукупності ОЕАН (визначеного з урахуванням рівня надійності результатів обстеження, який необхідно забезпечити, вартості обстеження та очікуваного рівня участі домогосподарств та осіб в обстеженні), до оцінки загальної кількості неінституційних домогосподарств в Україні (визначеної з використанням даних Всеукраїнського перепису населення, даних демографічної статистики, даних соціальної статистики, даних попередніх обсте-

жень ОЕАН та ОУЖД та ін.). Таким чином, базова вага домогосподарств в ОЕАН дорівнює $w_{Bi} = \frac{1}{f} \approx 554$.

Доцільно зазначити, що величина частки відбору домогосподарств f за своїм змістом суттєво відрізняється від величини загальної імовірності включення домогосподарства до вибірки p_i , хоча для самозваженої вибірки ОЕАН і приймається, що $p_i = f$. При розрахунку системи статистичних ваг ОЕАН використовується, насамперед, величина p_i та її складові. Це пояснюється тим, що в процесі побудови вибірки необхідно контролювати величини імовірностей відбору одиниць різних ступенів, оцінювати відхилення теоретичних імовірностей відбору домогосподарств від фактичних тощо.

2.2. Врахування фактичної імовірності відбору

Фактичні імовірності відбору домогосподарств в межах територіальних одиниць можуть відрізнятися від теоретичних, оскільки на практиці необхідно враховувати різницю в умовах роботи статистиків з обстежень. Наприклад, в ОЕАН прийнято, що у великих містах квартальне навантаження статистика з обстежень складає 72 домогосподарства, він обстежує дві ВТОВ, в кожній з яких відбирається по 36 домогосподарств, а у малих містах – 54 домогосподарства, статистик обстежує дві ВТОВ, в яких відбирається по 27 домогосподарств (див. рис. 2).

Тому необхідно скоригувати базові ваги домогосподарств з урахуванням фактичної імовірності відбору.

2.2.1. Розрахунок коефіцієнтів врахування фактичної імовірності відбору домогосподарств

Коефіцієнт урахування фактичної імовірності відбору i -го домогосподарства k_{fi} розраховується за формулою:

$$k_{fi} = \begin{cases} \frac{P_{3i}}{P_{fi}} & \text{по міських поселеннях} \\ \frac{P_{2i}}{P_{fi}} & \text{по сільській місцевості} \end{cases}, \quad (15)$$

де P_{fi} – фактична імовірність відбору домогосподарства в межах територіальної одиниці. Величина P_{fi} визначається за формулою:

$$P_{fi} = \frac{n_f}{N_f}, \quad (16)$$

де:

n_f – фактична кількість домогосподарств, що відібрані для обстеження у територіальній одиниці;

N_f – загальна кількість домогосподарств у територіальній одиниці.

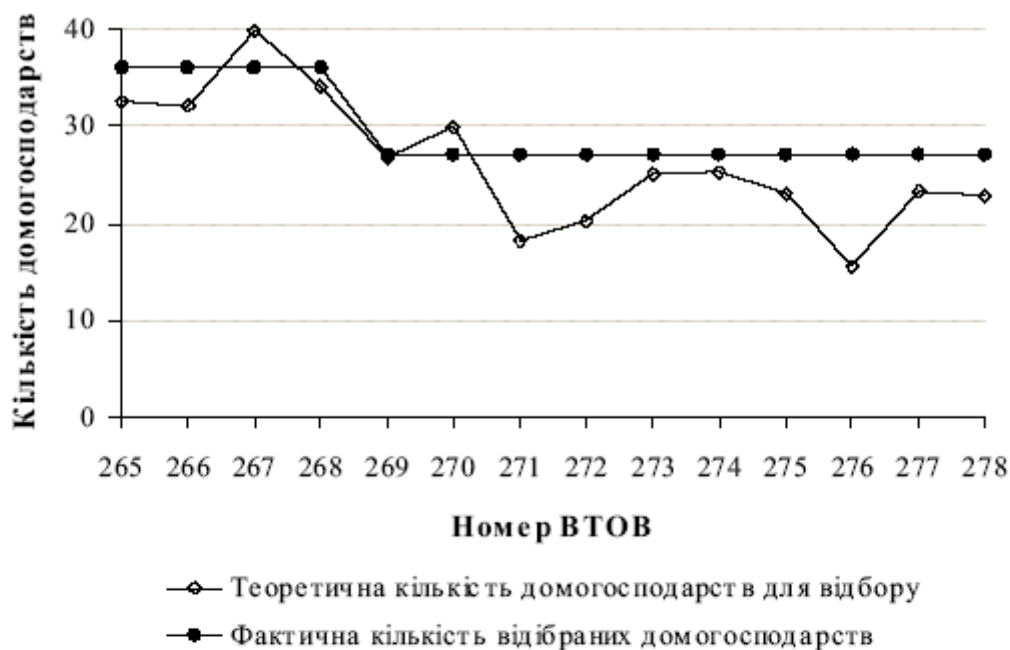


Рис. 2. Фактична кількість відібраних домогосподарств та теоретична кількість домогосподарств для відбору по міських ВТОВ Івано-Франківської області для ОЕАН у 2004-2008 рр.

2.2.2. Розрахунок ваг домогосподарств з урахуванням фактичних імовірностей відбору домогосподарств

Після визначення коефіцієнтів коригування ваг k_{li} розраховуються нові ваги домогосподарств w_{li} за формулою:

$$w_{li} = w_{Bi} \cdot k_{li} . \quad (17)$$

Статистичні ваги після врахування фактичних імовірностей відбору домогосподарств однакові в межах кожної окремої міської ВТОВ і кожної окремої сільської ПТОВ.

2.3. Врахування відмов домогосподарств від обстеження

Для врахування відмов домогосподарств від обстеження будується спеціальна система комірок. Система комірок – це система ознак для групування за ними одиниць генеральної та (або) вибіркової сукупностей. Головні завдання, що вирішуються при побудові системи комірок, це врахування при коригуванні статистичних ваг домогосподарств особливостей територіального охоплення вибірки і забезпечення певної кількості домогосподарств у кожній комірці.

При побудові системи комірок в ОЕАН використовуються наступні класифікаційні змінні:

- регіон – 27 дискретних значень (Автономна Республіка Крим, 24 області України, м. Київ, м. Севастополь);
- тип місцевості – 2 дискретних значення (міські поселення, сільська місцевість).

Крім того, враховується ступінь територіальної близькості груп домогосподарств, що включені до певної комірки.

Використання зазначеної вище інформації при побудові системи комірок реалізується шляхом об'єднання окремих територіальних одиниць за спеціальною схемою: відповідно ВТОВ у міських поселеннях і ПТОВ у сільській місцевості.

Встановлена мінімальна кількість домогосподарств, що були обстежені в кожній комірці, дорівнює 50. Якщо у деякій комірці кількість опитаних домогосподарств менша встановленої величини, ця комірка об'єднується з іншою (суміжною), що відноситься до тієї ж адміністративно-територіальної одиниці. Таким чином для кожного конкретного обстеження будується окрема система комірок для коригування статистичних ваг домогосподарств.

Загальна кількість домогосподарств у комірці n_l визначається за результатами проведення відбору домогосподарств по територіальних одиницях. При цьому з загальної кількості домогосподарств у комірці виключаються нежитлові приміщення, зачинені приміщення та неіснуючі адреси.

2.3.1. Розрахунок коефіцієнтів врахування відмов домогосподарств від обстеження

Коефіцієнт врахування відмов домогосподарств від обстеження k_{2i} розраховується по кожній l -ій комірці за формулою:

$$k_{2i} = \frac{\sum_{i=1}^n w_{1i} \cdot \lambda_{li}}{\sum_{i=1}^n w_{1i} \cdot \lambda_{li} \cdot \eta_{li}}, \quad (18)$$

де:

$$\lambda_{li} = \begin{cases} 1, & \text{якщо } i \in M_l \\ 0, & \text{якщо } i \notin M_l \end{cases};$$

M_l – множина домогосподарств l -ої комірки;

1, якщо домогосподарство взяло участь в обстеженні;

0, якщо домогосподарство відмовилось від обстеження.

Величини коефіцієнтів врахування відмов домогосподарств від обстеження k_{2i} однакові в межах кожної окремої комірки.

2.3.2. Розрахунок ваг домогосподарств з урахуванням відмов домогосподарств від обстеження

Ваги домогосподарств з урахуванням відмов домогосподарств від обстеження w_{2i} розраховуються за формулою:

$$w_{2i} = w_{1i} \cdot k_{2i} \quad (19)$$

2.4. Врахування відмов осіб від обстеження

Для врахування відмов від інтерв'ю окремих осіб в межах обстежених домогосподарств будується система комірок.

Код комірки визначається аналогічно коду комірки для врахування відмов домогосподарств від обстеження. Комірки нумеруються, починаючи з одиниці.

Загальна кількість осіб у комірці n_p визначається за результатами проведення інтерв'ю за поточний місяць.

2.4.1. Розрахунок коефіцієнтів урахування відмов осіб від обстеження

Коефіцієнт урахування відмов осіб від обстеження k_{3i} розраховується по кожній p -ій комірці за формулою:

$$k_{3i} = \frac{\sum_{i=1}^n w_{2i} \cdot \lambda_{pi}}{\sum_{i=1}^n w_{2i} \cdot \lambda_{pi} \cdot \eta_{pi}}, \quad (20)$$

де:

$$\lambda_{pi} = \begin{cases} 1, & \text{якщо } p \in M_p \\ 0, & \text{якщо } p \notin M_p \end{cases};$$

M_p – множина осіб p -ої комірки;

1, якщо особа взяла участь у обстеженні;

0, якщо особа відмовилась від участі в обстеженні;

Величини коефіцієнтів урахування відмов осіб від обстеження k_{3i} однакові в межах кожної окремої комірки.

2.4.2. Розрахунок ваг з урахуванням відмов осіб від інтерв'ю

Ваги осіб з урахуванням відмов від інтерв'ю w_{3i} розраховуються за формулою:

$$w_{3i} = w_{2i} \cdot k_{3i} \quad (21)$$

2.5. Усунення постстратифікаційних ефектів

Для усунення постстратифікаційних ефектів будується система комірок за такими класифікаційними змінними:

- регіон – 27 дискретних значень (25 областей, м. Київ та м. Севастополь);
- тип місцевості – 2 дискретних значення (міські поселення, сільська місцевість);
- стать респондента – 2 дискретних значення (жіноча, чоловіча);
- вікова група, до якої входить респондент – 11 дискретних значень (15-19 років, 20-24 роки, 25-29 років, 30-34 роки, 35-39 років, 40-44 роки, 45-49 років, 50-54 роки, 55-59 років, 60-64 роки, 65-70 років).

2.5.1. Визначення коду комірки c_q

Код комірки розраховується за формулою:

$$c_q = \text{ознака регіону} \cdot 1000 + \text{ознака типу місцевості} \cdot 100 + \\ + \text{ознака статі} \cdot 10 + \text{ознака вікової групи}$$

Комірки нумеруються, починаючи з одиниці після упорядкування їх за кодом.

Загальна кількість осіб у комірці n_q визначається за результатами проведення інтерв'ю за поточний місяць.

Загальна кількість осіб генеральної сукупності у комірці N_q визначається один раз на рік за даними демографічної статистики на 1 січня поточного року.

2.5.2. Розрахунок коефіцієнтів коригування ваг для усунення постстратифікаційних ефектів

Коефіцієнти для усунення постстратифікаційних ефектів k_{4i} розраховуються по кожній q -ій комірці за формулою:

$$k_{4q} = \frac{N_q}{\sum_r w_{3r} \cdot \lambda_{qr}}, \quad (22)$$

де:

$$\lambda_{qr} = \begin{cases} 1, & \text{якщо } q \in M_q \\ 0, & \text{якщо } q \notin M_q \end{cases};$$

M_q – множина осіб q -ої комірки.

Величини коефіцієнтів урахування відмов осіб від обстеження k_{4i} однакові в межах кожної окремої комірки.

Ілюстративні результати розрахунку коефіцієнтів для усунення постстратифікаційних ефектів наведено у додатку А.

2.5.3. Розрахунок ваг осіб з урахуванням коефіцієнтів для усунення постстратифікаційних ефектів

Ваги осіб після усунення постстратифікаційних ефектів w_{4i} розраховуються за формулою:

$$w_{4i} = w_{3i} \cdot k_{4i} \quad (23)$$

3. Об'єднання даних, отриманих за основною вибіркою та за вибіркою обстеження сільськогосподарської діяльності домогосподарств

Однією з відмінностей побудови вибірки ОЕАН для обстежень 2004-2008 років у порівнянні з вибіркою, діючою у 1999-2003 роках, є те, що у сільській місцевості вибірка охоплює всі сільські райони. До цього обстежувались лише 189 випадково відібраних районів.

Формування вибірки домогосподарств у сільській місцевості здійснюється за рахунок двох складових – вибірки безпосередньо для ОЕАН та вибірки для обстеження домогосподарств з питань їх сільськогосподарської діяльності.

Тому виникає необхідність в об'єднанні даних ОЕАН, отриманих на базі основної вибірки та вибірки обстеження сільськогосподарської діяльності домогосподарств (ОСГД). Процедура здійснюється із застосуванням методу приєднання даних на мікрорівні, тобто на рівні осіб. Цей підхід використовується в умовах, коли вибіркові дані по одиницях однакового рівня (членів домогосподарств віком 15-70 років) та за однаковими ознаками (за ідентичними анкетами) одержуються на основі різних методів побудови вибірки. При побудові вибірки обстеження сільськогосподарської діяльності домогосподарств відбір сільських рад здійснюється пропорційно площі земель.

Схема процедури приєднання даних наведена на рис. 3.

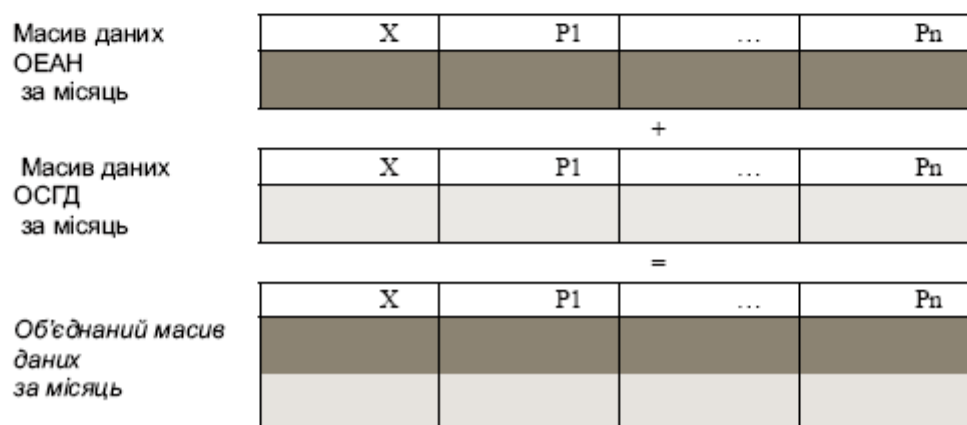


Рис. 3. Схема об'єднання щомісячних даних ОЕАН за основною вибіркою та за вибіркою ОСГД в єдиний масив

Результуючі ваги осіб, розраховані окремо по кожному з масивів, коригуються в межах сільської місцевості кожного регіону з використанням $\hat{\phi}_{ЗН}$ коефіцієнтів для зайнятого населення та $\hat{\phi}_{БР}$ для безробітного населення [7]. Оцінки коефіцієнтів $\hat{\phi}_{ЗН}$ і $\hat{\phi}_{БР}$ визначаються з умов мінімізації загальної похибки оцінювання показників чисельності зайнятого та безробітного населення за об'єднаним масивом даних:

$$\begin{aligned} \hat{\phi}_{ЗН} &= \frac{SE^2(\hat{Y}_{ЗН}^{(ОСГД)})}{SE^2(\hat{Y}_{ЗН}^{(ОЕАН)}) + SE^2(\hat{Y}_{ЗН}^{(ОСГД)})} \text{ для зайнятого населення} \\ \hat{\phi}_{БР} &= \frac{SE^2(\hat{Y}_{БР}^{(ОСГД)})}{SE^2(\hat{Y}_{БР}^{(ОЕАН)}) + SE^2(\hat{Y}_{БР}^{(ОСГД)})} \text{ для безробітного населення,} \end{aligned} \quad (24)$$

де:

$SE(\hat{Y}_{ЗН}^{(ОЕАН)})$ – стандартна похибка оцінки чисельності зайнятого населення $\hat{Y}_{ЗН}^{(ОЕАН)}$ за вибіркою ОЕАН;

$SE(\hat{Y}_{ЗН}^{(ОСГД)})$ – стандартна похибка оцінки чисельності зайнятого населення $\hat{Y}_{ЗН}^{(ОСГД)}$ за вибіркою ОСГД;

$SE(\hat{Y}_{БР}^{(ОЕАН)})$ – стандартна похибка оцінки чисельності безробітного населення $\hat{Y}_{БР}^{(ОЕАН)}$ за вибіркою ОЕАН;

$SE(\hat{Y}_{БР}^{(ОСГД)})$ – стандартна похибка оцінки чисельності безробітного населення $\hat{Y}_{БР}^{(ОСГД)}$ за вибіркою ОСГД.

Коефіцієнти $\hat{\phi}_{ЗН}$ і $\hat{\phi}_{БР}$ розраховуються щомісяця по кожному регіону окремо. У табл. 1 для ілюстрації наведені значення коефіцієнтів $\hat{\phi}_{ЗН}$ і $\hat{\phi}_{БР}$ для місяців першого кварталу 2005 року.

Коефіцієнт коригування системи статистичних ваг k_{5i} розраховується в межах кожного регіону у два етапи.

На першому етапі велич k_{5i} розраховується для зайнятих та безробітних осіб сільської місцевості за формулою:

$$\begin{aligned} k_{5i} &= \frac{\hat{\phi}_{ЗН} \text{ для зайнятих осіб за вибіркою ОЕАН}}{(1 - \hat{\phi}_{ЗН}) \text{ для зайнятих осіб за вибіркою ОСГД}} \\ &\quad \frac{\hat{\phi}_{БР} \text{ для безробітних осіб за вибіркою ОЕАН}}{(1 - \hat{\phi}_{БР}) \text{ для безробітних осіб за вибіркою ОСГД}} \end{aligned} \quad (25)$$

Розраховуються скориговані статистичні ваги зайнятих і безробітних осіб по сільській місцевості кожного регіону за формулою:

$$w_{5i} = w_{4i} \cdot k_{5i} \quad (26)$$

Таблиця 1

Значення коефіцієнтів $\hat{\phi}_{3H}$ і $\hat{\phi}_{BP}$ при об'єднанні даних
ОЕАН та ОСГД у січні, лютому та березні 2005 року

Регіон	$\hat{\phi}_{3H}$			$\hat{\phi}_{BP}$		
	січень	лютий	березень	січень	лютий	березень
АР Крим	0,38	0,37	0,40	0,32	0,16	0,09
Вінницька область	0,43	0,42	0,43	0,37	0,32	0,23
Волинська область	0,49	0,52	0,51	0,36	0,50	0,48
Дніпропетровська область	0,41	0,43	0,44	0,14	0,24	0,44
Донецька область	0,39	0,37	0,39	0,25	0,35	0,38
Житомирська область	0,48	0,46	0,46	0,42	0,39	0,50
Закарпатська область	0,43	0,41	0,39	0,47	0,24	0,27
Запорізька область	0,44	0,41	0,44	0,68	0,30	0,30
Івано-Франківська область	0,48	0,47	0,45	0,28	0,36	0,50
Київська область	0,41	0,41	0,40	0,71	0,45	0,48
Кіровоградська область	0,43	0,43	0,42	0,24	0,53	0,54
Луганська область	0,39	0,39	0,45	0,45	0,14	0,16
Львівська область	0,46	0,46	0,45	0,45	0,46	0,44
Миколаївська область	0,48	0,45	0,45	0,66	0,68	0,76
Одеська область	0,43	0,43	0,41	0,37	0,19	0,37
Полтавська область	0,43	0,46	0,45	0,25	0,28	0,29
Рівненська область	0,46	0,46	0,46	0,39	0,45	0,49
Сумська область	0,40	0,40	0,38	0,26	0,31	0,31
Тернопільська область	0,45	0,47	0,47	0,34	0,59	0,59
Харківська область	0,46	0,46	0,44	0,49	0,61	0,29
Херсонська область	0,49	0,50	0,47	0,70	0,83	0,37
Хмельницька область	0,43	0,43	0,41	0,23	0,40	0,32
Черкаська область	0,41	0,40	0,41	0,21	0,23	0,28
Чернівецька область	0,49	0,46	0,49	0,17	0,36	0,37
Чернігівська область	0,43	0,43	0,43	0,36	0,40	0,25

На другому етапі величина k_{5i} розраховується для економічно неактивних осіб сільської місцевості кожного регіону за формулою:

$$k_{5i} = \frac{N_{15-70} \left(\hat{Y}_{3H}^{(OЕАН)} \hat{Y}_{3H}^{(ОСГД)} \hat{Y}_{BP}^{(OЕАН)} \hat{Y}_{BP}^{(ОСГД)} \right)}{\hat{Y}_{EH}^{(OЕАН)} \hat{Y}_{EH}^{(ОСГД)}}, \quad (27)$$

де:

N_{15-70} – загальна чисельність працездатного населення по сільській місцевості регіону розрахована за зовнішніми даними;

$\hat{Y}_{3H}^{(OЕАН)}$ – оцінка чисельності зайнятого населення за вибіркою ОЕАН з урахуванням скоригованих статистичних ваг w_{5i} ;

$\hat{Y}_{3H}^{(ОСГД)}$ – оцінка чисельності зайнятого населення за вибіркою ОСГД з урахуванням скоригованих статистичних ваг w_{5i} ;

$\hat{Y}_{BP}^{(OEAN)}$ – оцінка чисельності безробітного населення за вибіркою OEAN з урахуванням скоригованих статистичних ваг w_{5i} ;

$\hat{Y}_{BP}^{(OCГД)}$ – оцінка чисельності безробітного населення за вибіркою OCГД з урахуванням скоригованих статистичних ваг w_{5i} ;

$\hat{Y}_{EH}^{(OEAN)}$ – оцінка чисельності економічно неактивного населення за вибіркою OEAN з урахуванням статистичних ваг w_{4i} ;

$\hat{Y}_{EH}^{(OCГД)}$ – оцінка чисельності економічно неактивного населення за вибіркою OCГД з урахуванням статистичних ваг w_{4i} .

Коригування системи статистичних ваг економічно неактивного населення сільської місцевості за регіонами у об'єднаному масиві даних здійснюється за формулою аналогічною (26).

Слід наголосити, що описана у цьому підрозділі процедура застосовується лише для одиниць вибірки сільської місцевості.

При формуванні результуючого масиву після зазначеного коригування статистичних ваг необхідно здійснити процедуру постстратифікації для всього масиву (див. підрозділ 2.5).

Для міських одиниць вибірки остаточною будуть ваги (23), отримані після усунення постстратифікаційних ефектів.

4. Порядок визначення якості системи ваг

Якість системи ваг OEAN визначається після кожного етапу їх коригування. При цьому розраховуються наступні характеристики.

1) Середня вага по всіх домогосподарствах або особах (в залежності від етапу розрахунку системи ваг), як окремий множник – коефіцієнт інфляції F [7]. Величина $F_{(j)}$ для j -го етапу розрахунку ваг визначається за формулою:

$$F_{(j)} = \frac{\sum_{i=1}^n w_{(j)i}}{n} \quad (28)$$

2) Максимальне значення $w_{(j)\max}$ та мінімальне значення $w_{(j)\min}$ ваги домогосподарств (осіб). Розраховуються відношення, $R_{(j)1}$, $R_{(j)2}$ та $R_{(j)3}$ за формулами:

$$R_{(j)1} = \frac{F_{(j)}}{w_{(j)\min}} \quad R_{(j)2} = \frac{w_{(j)\max}}{F_{(j)}} \quad R_{(j)3} = \frac{w_{(j)\max}}{w_{(j)\min}} \quad (29)$$

При аналізі отриманих результатів мати на увазі, що є неприпустимим значення $w_{\min}^{(I)} = 0$.

3) Розмах варіації ваг за формулою:

$$\Delta R_{(j)} = w_{(j)\max} - w_{(j)\min} \quad (30)$$

4) Відносна вага (відношення ваги домогосподарства (особи) до коефіцієнта інфляції) для кожного домогосподарства або особи. Ця величина приймає значення, близькою до одиниці (в середньому по всіх домогосподарствах (особах) відносна вага дорівнює одиниці).

Використання відносних (приведених) ваг дозволяє більш наглядно порівнювати незважені дані та дані після кожного етапу зважування, порівнювати дис-

персії ваг на кожному етапі їх коригування, виявляти тенденції у зміні максимальних та мінімальних значень тощо. Крім того, порівняння суміжних коефіцієнтів інфляції дозволяє виявити етапи коригування ваг, які в середньому мають найбільший вплив на ваги домогосподарств (осіб).

Після кожного етапу коригування ваг слід перевіряти відносні ваги на наявність дуже малих (≈ 0) та від'ємних значень, розраховувати відношення максимальної відносної ваги до мінімальної.

4) Відносне зростання дисперсії вибірових оцінок, джерелом яких є варіативність ваг по одиницях вибірки, k_L . Величина k_L визначається за формулою [8]:

$$k_L = \frac{n \sum_{i=1}^n (w_i)^2}{(\sum_{i=1}^n w_i)^2} \quad (31)$$

Якщо дисперсія вибірових оцінок ($\mathcal{V}$) показника $\bar{Y}$ без урахування ваг визначається величиною $V(\bar{Y})$, то за рахунок власної варіативності ваг дисперсія зваженої оцінки буде визначатися величиною $k_L \cdot V(\bar{Y})$, що еквівалентно зменшенню обсягу вибірки у k_L разів.

$$CV(w_{(j)}) = \sqrt{\frac{1}{n \cdot F_{(j)}^2} \cdot \sum_{i=1}^n (w_{(j)i} - F_{(j)})^2}, \quad (32)$$

де w_i – вага i -го домогосподарства (особи), $i=1, 2, \dots, n$;

6) Відношення коефіцієнтів інфляції для поточного та попереднього етапів зважування на j -му етапі $R_{F(j)}$:

$$R_{F(j)} = \frac{F_{(j)}}{F_{(j-1)}} \quad (33)$$

7) Відношення коефіцієнтів варіації ваг поточного та попереднього етапів зважування $R_{(j)4}$ за формулою:

$$R_{(j)4} = \frac{CV(w_{(j)})}{CV(w_{(j-1)})} \quad (34)$$

8) Коефіцієнт кореляції Пірсона $r_{(j,j-1)}$ для суміжних векторів ваг за формулою:

$$r_{(j,j-1)} = \frac{\sum_{i=1}^n (w_{(j-1)i} - \bar{w}_{(j-1)}) \cdot (w_{(j)i} - \bar{w}_{(j)})}{\sqrt{\sum_{i=1}^n (w_{(j-1)i} - \bar{w}_{(j-1)})^2 \cdot \sum_{i=1}^n (w_{(j)i} - \bar{w}_{(j)})^2}} \quad (35)$$

Коефіцієнт кореляції характеризує щільність лінійного статистичного зв'язку системи ваг, що відповідають суміжним етапам розрахунку ваг. Чим ближчим є модуль величини цього коефіцієнту до одиниці, тим менше розподіл поточних ваг відхиляється від розподілу попереднього етапу і, в цілому, від початкових ваг дизайну.

9) Розраховується оцінка загальної кількості домогосподарств у генеральній сукупності $N_{(j)}$ за формулою:

$$N_{(j)} = \sum_{i=1}^n w_{(j)i} \quad (36)$$

Оцінка загальної кількості домогосподарств $N_{(j)}$ порівнюється з оцінкою загальної кількості домогосподарств з зовнішніх джерел N_{3I} (попередньою оцінкою загальної кількості домогосподарств, що використовувалась при побудові вибірки, даними демографічної статистики та обстеження умов життя домогосподарств або з іншою (достовірною) інформацією щодо загальної кількості неінституційних домогосподарств в Україні). Розраховується відносне відхилення за формулою:

$$D_{(j)} = \frac{N_{3I} - N_{(j)}}{N_{3I}} \cdot 100\% \quad (37)$$

Отримані результати аналізуються на підставі існуючих даних щодо рівня репрезентативності вибірки.

Якість розрахованої системи ваг визначається на основі отриманих значень характеристик шляхом їх порівняння із встановленими граничними значеннями, характеристиками якості за попередні періоди обстежень та на основі експертних оцінок. При цьому слід враховувати, що рівень якості системи ваг залежить від результатів обстежень і тому для кожного окремого обстеження висновок щодо якості системи ваг необхідно обґрунтовувати окремо.

10) По масиву первинних даних для осіб, що проживають в домогосподарствах, по яких проведено інтерв'ю або є обов'язкова інформація, з використанням ваг $w_{(j)i}$, розраховуються оцінки загальної чисельності населення України та чисельності населення за розрізами регіонів – тип місцевості, стать, вікові групи. Результати порівнюються з даними демографічної статистики.

Приклад окремих результатів оцінювання якості ваг наведено у табл.2. Наведені дані ілюструють зміну окремих характеристик якості системи ваг ОЕАН протягом 1999-2004 років.

Таблиця 2

**Зміна окремих характеристик якості системи ваг
ОЕАН за 1999-2004 роки**

Характеристики	Рік					
	1999	2000	2001	2002	2003	2004
Середня вага, F	743	738	658	647	653	630
Мінімальна вага, w_{min}	185	192	196	187	195	66
Максимальна вага, w_{max}	2433	3028	2752	2079	2665	4344
Коефіцієнт варіації ваг, $CV_w, \%$	36,48	37,75	38,56	38,66	41,42	45,73
$R_1 = \frac{\overline{w}}{w_{min}}$	4,02	3,84	3,36	3,45	3,35	9,54
$R_2 = \frac{w_{max}}{\overline{w}}$	3,28	4,10	4,18	3,21	4,08	6,89
$R_3 = \frac{w_{max}}{w_{min}}$	13,15	15,74	14,04	11,11	13,65	65,71
Розмах варіації ваг, ΔR	2248	2835	2556	1892	2470	4278

Як видно з наведених даних, спостерігається тенденція до зменшення величини середньої ваги. Враховуючи, що обсяг квартальної вибірки, на базі якої проводилось обстеження протягом 1999–2004 років, практично не змінювався, це можна пояснити зміною дизайну вибірки, зменшенням кількості домогосподарств у генеральній сукупності та зусиллями фахівців Держкомстату щодо зниження рівня відмов домогосподарств від обстеження. В обстеженні також збільшується коефіцієнт варіації ваг, але несуттєво. Величини інших характеристик в ОЕАН свідчать про погіршення якості системи ваг, особливо у 2004 році. Це пояснюється ефектами „старіння” вибірки до 2003 року, а у 2004 – впровадженням нової системи організації державних вибіркових обстежень населення [3].

5. Інформаційна база розрахунку системи статистичних ваг ОЕАН

Інформаційна база для реалізації процедур розрахунку системи ваг за результатами поточного щомісячного обстеження складається з трьох основних частин:

- характеристик вибіркової сукупності домогосподарств;
- результатів ОЕАН;
- інформації з зовнішніх джерел.

5.1. Характеристики вибіркової сукупності домогосподарств

Характеристики вибіркової сукупності включають дані щодо дизайну та основних параметрів вибірки, а саме:

- загальнодержавний номер територіальної одиниці (ВТОВ – по міських поселеннях, ПТОВ – по сільській місцевості);
- імовірність відбору одиниць першого ступеня (міських поселень та сільських рад), P_{1i} ;
- імовірність відбору одиниць другого ступеня (інструкторські ділянки – по міських поселеннях, домогосподарства – по сільській місцевості), P_{2i} ;
- імовірність відбору домогосподарств по міських поселеннях, P_{3i} ;
- загальна кількість домогосподарств у ВТОВ по міських поселеннях, N_{2i} ;
- загальна кількість домогосподарств у ПТОВ по сільській місцевості, N_{1i} ;
- фактична кількість відібраних для обстеження домогосподарств у територіальній одиниці вибірки (у ВТОВ – по міських поселеннях, у ПТОВ – по сільській місцевості), n_{TO} .

Характеристики вибіркової сукупності залишаються незмінними протягом всього терміну дії територіальної вибірки ОЕАН (для поточного циклу обстежень – це 2004–2008 роки). Разом з тим у випадку вимушеної заміни територіальних одиниць вибірки або зміни фактичних кількостей відібраних домогосподарств необхідно скоригувати відповідні параметри вибірки.

5.2. Результати обстеження

Результати ОЕАН включають наступні дані:

- масиви обстежених домогосподарств та осіб за результатами поточного ОЕАН:

- код домогосподарства;
- код особи;
- стать особи;
- вік особи;

-
- ознака належності до вибірки ОСГД;
 - статистична вага особи в ОСГД;
 - статус економічної активності особи;
 - загальнодержавний номер територіальної одиниці (ВТОВ по міських поселеннях, ПТОВ по сільській місцевості);
 - код регіону;
 - тип місцевості.
- номер комірки для врахування відмов домогосподарств від обстеження, l ;
 - код комірки, c_p ;
 - загальна початкова кількість домогосподарств у комірці, n_p ;
 - номер комірки для врахування відмов осіб від обстеження, p ;
 - код комірки, c_q ;
 - загальна початкова кількість осіб у комірці, n_p ;
 - код комірки, c_q ;
 - загальна початкова кількість осіб у комірці, n_q .

5.3. Додаткові дані

Інформація з зовнішніх джерел містить дані демографічної статистики щодо статеві-вікової структури населення за п'ятирічними віковими групами на початок поточного року, що публікуються у статистичному збірнику „Склад населення України за статтю та віком”, а також дані про кількість неінституційних домогосподарств.

Масив зовнішньої інформації створюється один раз на рік перед розрахунком системи статистичних ваг домогосподарств за результатами ОЕАН у січні поточного року.

ЗАКЛЮЧНІ ПОЛОЖЕННЯ

Методика поширення результатів ОЕАН на генеральну сукупність передбачає розрахунок системи ваг на мікрорівні, тобто для кожного обстеженого домогосподарства та особи, на щомісячній основі. Таким чином, можливість реалізації даної методики на практиці вимагає застосування автоматизованих розрахункових процедур при побудові системи ваг, розрахунках характеристик їх якості, оцінювання показників тощо. Технологія розрахунку системи ваг повинна бути часткою загальної технології обробки щомісячних даних ОЕАН.

Методика може бути удосконалена в напрямку більш детального відображення особливостей реалізації ОЕАН, зокрема, для врахування схеми ротації домогосподарств у вибірці, більш ефективного використання наявних додаткових даних шляхом розробки та реалізації процедур калібрації системи ваг, або у напрямку забезпечення можливості застосування методів оцінювання показників на основі моделей та ін.

ЗАТВЕРДЖЕНО

Наказ Держкомстату України
від 14.11.2006 р. № 520

Література

1. Вибіркове спостереження: Термінологічний словник. – К.: НТК статистичних досліджень, 2004.
2. Методологічні основи формування вибірових сукупностей для проведення органами державної статистики України базових державних вибірових обстежень населення (домогосподарств) / Затверджені наказом Держкомстату України від 2 серпня 2005 р., №223. – К., 2005 р.
3. Економічна активність населення України у 2004 році: Стат. збірник. – К.: Державний комітет статистики України, 2005.
4. Kish L. Survey sampling. – New York: Wiley, 1995.
5. Kalton G. Introduction to Survey Sampling. – Beverly Hills: Sage, 1983.
6. Workshop on Labour Force Surveys for CIS Countries and the Baltic States run by the International Labour Office, Bureau of Statistics hosted by State Statistics Committee of Ukraine (Kyiv, 14-19 September 1998). Part Three: sample design, data evaluation, field operations (Vijay Verma).
7. Сариогло В.Г. Проблеми статистичного зважування вибірових даних. – К.: ІВЦ Держкомстату України, 2005.
8. Kish L.: Selected papers. – New York: Wiley, 2003.
9. SPSS® Base 8.0 для Windows®, Applications Guide, Copyright© 1998 СПСС Русь.

Додаток А

РЕЗУЛЬТАТИ РОЗРАХУНКУ КОЕФІЦІЄНТІВ ДЛЯ УСУНЕННЯ ПОСТСТРАТИФІКАЦІЙНИХ ЕФЕКТІВ АВТНОМНА РЕСПУБЛІКА КРИМ (ОЕАН, січень 2005)

	Вікові групи	Зовнішня інформація	Фактична чисельність	Коефіцієнт	Вікові групи	Зовнішня інформація	Фактична чисельність	Коефіцієнт
Жінки	Міська місцевість				Сільська місцевість			
	15-19	48359	39913	1,2116	15-19	31451	17188	1,8299
	20-24	49300	48486	1,0168	20-24	29752	27218	1,0931
	23-29	45711	45711	1,0000	23-29	25643	10500	2,4422
	30-34	44420	34854	1,2745	30-34	23622	19186	1,2312
	35-39	44822	44822	1,0000	35-39	23912	23193	1,0310
	40-44	54617	54617	1,0000	40-44	29979	16853	1,7788
	45-49	56577	56422	1,0027	45-49	30740	27613	1,1133
	50-54	55527	55527	1,0000	50-54	27833	23902	1,1645
	55-59	43578	43578	1,0000	55-59	20015	19158	1,0447
	60-64	34267	34267	1,0000	60-64	20953	18683	1,1215
	65-70	51934	51934	1,0000	65-70	31039	31039	1,0000
Чоловіки	15-19	49213	47549	1,0350	15-19	33636	29695	1,1327
	20-24	52348	49495	1,0576	20-24	30656	14000	2,1897
	23-29	44440	35886	1,2384	23-29	27143	3500	7,7551
	30-34	41020	41020	1,0000	30-34	23774	16486	1,4420
	35-39	38368	36245	1,0586	35-39	23908	22246	1,0747
	40-44	45416	45008	1,0091	40-44	29537	22704	1,3010
	45-49	44238	44238	1,0000	45-49	29428	19817	1,4850
	50-54	42150	42150	1,0000	50-54	25301	23813	1,0625
	55-59	31407	31407	1,0000	55-59	15638	15638	1,0000
	60-64	23310	23310	1,0000	60-64	14148	13388	1,0568
	65-70	31595	31595	1,0000	65-70	20259	20259	1,0000