

МЕТОДИКА РОЗРАХУНКУ СИСТЕМИ СТАТИСТИЧНИХ ВАГ ДЛЯ ПОШИРЕННЯ РЕЗУЛЬТАТІВ ДЕРЖАВНОГО ВИБІРКОВОГО ОБСТЕЖЕННЯ УМОВ ЖИТТЯ ДОМОГОСПОДАРСТВ НА ГЕНЕРАЛЬНУ СУКУПНІСТЬ

Анотація

Методика визначає основні принципи та порядок розрахунку системи статистичних ваг домогосподарств на державному вибірковому обстеженні умов життя домогосподарств. Система ваг забезпечує можливість поширення результатів вибіркового обстеження на генеральну сукупність – всі неінституційні домогосподарства України та підвищення якості вимірювання показників шляхом урахування фактичних імовірностей включення одиниць генеральної сукупності до вибірки, рівнів відмов від обстеження по групах одиниць спостереження, достовірних даних із зовнішніх джерел тощо.

Цей документ містить опис принципів екстраполяції (поширення) даних вибіркового обстеження умов життя домогосподарств на генеральну сукупність, порядок розрахунку складових системи ваг та визначення її якості.

Методика розроблена на основі сучасних методологічних засад, що використовуються в міжнародній практиці проведення державних вибіркового обстежень населення, насамперед, в частині принципів забезпечення репрезентативності вибірки, методів урахування рівнів участі домогосподарств в обстеженні та методів узгодження результатів обстежень з наявною додатковою (зовнішньою) інформацією. Ця методика є однією з основних складових методичного забезпечення проведення вибіркового обстеження умов життя домогосподарств у 2004–2008 роках. Процедури та алгоритми, реалізація яких передбачена методикою, орієнтовані на використання сучасних комп'ютерних технологій обробки статистичної інформації.

Підготовка цієї методики була здійснена за підтримки проекту „Допомога Державному комітету статистики України” Міністерства у справах міжнародного розвитку Сполученого Королівства Великої Британії та Північної Ірландії.

Методичні матеріали призначені для використання працівниками органів державної статистики на державному та регіональному рівнях, міністерствами, іншими органами виконавчої влади, науковими організаціями, навчальними закладами та іншими користувачами статистичної інформації.

Дану редакцію Методики підготовлено фахівцями Інституту демографії та соціальних досліджень НАН України В.Г.Сариогло, О.М.Гладуном, О.В.Лисою, за участю директора департаменту обстежень домогосподарств Держкомстату України І.І. Осипової та за консультативної допомоги п. Пола Сміта (Paul Smith) Голови центру дизайну вибірки та оцінювання Національного офісу статистики Сполученого Королівства.

Передмова

Можливість екстраполяції (поширення) результатів вибіркового обстеження на генеральну сукупність – це один з головних факторів, що обумовлює ефективність вибіркового методів спостереження. За останні роки застосування цих методів набуло значного розвитку в галузях спостережень за рівнем та умовами життя домогосподарств, економічною діяльністю населення та ряду інших досліджень, що проводяться органами державної статистики України. Ці методи дозволяють за результатами дослідження певної частини генеральної сукупності оцінювати цільові показники для національного рівня, для окремих регіонів та різних верств населення.

Основним методологічним принципом, що визначає можливість екстраполяції результатів вибіркового обстеження на генеральну сукупність, є принцип репрезентативності, який вимагає незміщеності оцінок показників генеральної сукупності за результатами вибіркового обстеження, тобто їх наближеності до дійсних значень показників. Принцип репрезентативності для вибіркового обстежень реалізується через принцип імовірності вибірки, за яким кожний елемент генеральної сукупності повинен мати відому позитивну імовірність потрапити до вибірки при її формуванні. Можливість дотримання принципу репрезентативності в обстеженні умов життя домогосподарств (надалі ОУЖД) ускладнюється тим, що це обстеження має складний дизайн вибірки, характеризується значними і неоднорідними за регіонами та групами населення рівнями відмов від обстеження, вимагає застосування складних методів обробки даних, оцінювання показників та узгодження результатів обстеження з наявною зовнішньою інформацією. Для репрезентативності результатів ОУЖД є необхідною реалізація спеціальної багатоступеневої процедури їх статистичного зважування. Система статистичних ваг одиниць спостереження забезпечує адекватне врахування при оцінюванні показників реального дизайну вибірки, фактичних рівнів участі респондентів в обстеженні, наявних додаткових, зовнішніх по відношенню до результатів обстежень, даних. Система ваг є невід'ємною складовою результатів ОУЖД.

Методика розроблена з використанням положень „Методики зважування та перезважування результатів обстеження умов життя домогосподарств”, затвердженої наказом Державного комітету статистики України №88 від 17 березня 2000 року та „Методики взаємоузгодження ваг рівня домогосподарств і рівня осіб”, затвердженої наказом Державного комітету статистики України №5 від 4 січня 2002 року. Необхідність розробки нової методики обумовлена впровадженням у 2004 році нової системи проведення державних вибіркового обстежень населення у 2004–2008 рр., яка передбачає застосування зміненого, порівняно з системою обстежень 1999–2003 рр., дизайну вибірки ОУЖД, удосконалення процедур та технології розрахунку системи статистичних ваг та ін.

Визначення термінів [1]

Відбір з ймовірністю пропорційною розміру – спосіб відбору, при якому ймовірність включення одиниці до вибірки пропорційна значенню певної характеристики (розміру) одиниці.

Відбір систематичний – спосіб відбору, при якому зі списку одиниць генеральної сукупності відбирають одиниці спостереження з використанням встановленого кроку відбору (інтервалу, через який слід брати для спостереження одиниці (серії) залежно від кількості одиниць (серій), що необхідно відібрати). Початок відбору

визначають або з використанням генератору випадкових чисел із номерів одиниць спостереження, які відповідають першому інтервалу, або способом жеребкування, або шляхом розрахунку (наприклад, шляхом ділення кроку відбору на два (якщо одиниці генеральної сукупності у списку не ранжовані за ознакою).

Відбір імовірнісний – відбір, оснований на об'єктивних правилах випадкового відбору, під час якого кожна одиниця сукупності має відомий ненульовий шанс бути включеним до вибірки.

Відбір багатоступеневий – відбір, який відбувається у кілька послідовних ступенів (етапів): на першому ступені з основи вибірки відбираються певні відносно великі одиниці; на другому ступені всередині кожної відбраної на першому ступені одиниці формується своя основа вибірки, з якої відбираються свої одиниці, і так далі за кількістю ступенів відбору.

Вибіркова сукупність (вибірка) – сукупність одиниць спостереження, відібраних із генеральної сукупності за певними правилами, які забезпечують репрезентативність вибірки, для проведення статистичного спостереження.

Вибіркове обстеження населення (домогосподарств) – науково-обґрунтований вид несущільного спостереження, при якому обстежуються не всі одиниці сукупності, а лише певним чином відібрана їх частина, а одиницею спостереження є особа чи домогосподарство.

Вибірковий метод спостереження – науково-обґрунтована система правил відбору одиниць і способів характеристики сукупності досліджуваних одиниць, що вивчають. Вибірковий метод спостереження дає змогу розповсюдити висновки, отримані на основі вивчення частини сукупності (вибірки), на всю сукупність (генеральну).

Вторинна територіальна одиниця вибірки (ВТОВ) – територіальна одиниця, яка відібрана на другому ступені формування територіальної вибірки при застосуванні багатоступеневого територіального відбору.

Генеральна сукупність – множина одиниць, характеристики яких необхідно дослідити, і щодо яких буде зроблено узагальнення за результатами вибіркового обстеження.

Дизайн (план) вибірки – організаційно-логічна модель структури вибіркової сукупності та принципи її формування. Якістю плану вибірки в значній мірі визначається репрезентативність вибіркової сукупності, отриманої за визначеними у плані правилами.

Ймовірність – число, яке характеризує ступінь можливості настання певної події. Розраховується як відношення кількості випадків настання події до всіх можливих варіантів.

Калібрація – трансформація (коригування) системи ваг для одиниць спостереження, які прийняли участь у вибіркового обстеженні, шляхом мінімізації функції відстані між калібраційними вагами та вагами, що коригуються, за умови, що калібраційні ваги дають точні оцінки зовнішніх змінних, відомих для генеральної сукупності або її частин.

Обсяг вибірки – кількість одиниць у вибірковій сукупності. Необхідний для забезпечення репрезентативності та належної якості результатів вибіркового обстеження обсяг вибірки розраховують в залежності від способу відбору та типу вибірки. Величина обсягу вибірки залежить від варіації досліджуваних ознак та припустимої похибки вибірки.

Одиниця відбору – одиниця генеральної сукупності чи їх група, або одиниця основи вибірки, яка має ймовірність бути включеною до вибірки при реалізації процедури відбору.

Одиниця вибірки – одиниця відбору, яка увійшла до складу вибірки.

Основа вибірки – упорядкований певним чином перелік, список одиниць відбору. При використанні багатоступеневого відбору на кожному ступені формується своя основа вибірки.

Одиниця спостереження – первинна одиниця, від якої мають бути одержані необхідні статистичні дані. Вибір одиниці спостереження залежить від мети та конкретних умов проведення спостереження.

Первинна територіальна одиниця вибірки (ПТОВ) – територіальна одиниця, яка відібрана на першому ступені формування територіальної вибірки при застосуванні багатоступеневого територіального відбору.

Репрезентативність – властивість вибіркової сукупності повно та адекватно відображати основні характеристики генеральної сукупності.

Статистична вага – 1) кількість одиниць генеральної сукупності, які репрезентує (представляє) одна одиниця вибірки; 2) величина, обернена до ймовірності відбору одиниці вибірки.

Стратифікація – поділ генеральної сукупності (або її частини) на групи, однорідні за певною ознакою.

Територіальна вибірка – відібрані за певною процедурою територіальні одиниці, в яких проводиться вибіркоче обстеження.

Територіальна одиниця – адміністративно-територіальна одиниця України (наприклад: міста, селища міського типу, райони, сільради, тощо) або локальна територія, створена в межах адміністративно-територіальної одиниці (наприклад: переписні відділи, інструкторські дільниці, створені для проведення Всеукраїнського перепису населення; виборчі дільниці; території, що обслуговуються поштовими відділеннями, тощо).

[1] При підготовці цього підрозділу використані матеріали, викладені у термінологічному словнику "Вибіркове спостереження" [1].

1. ПРИНЦИПИ ЕКСТРАПОЛЯЦІЇ РЕЗУЛЬТАТІВ ВИБІРКОВОГО ОБСТЕЖЕННЯ УМОВ ЖИТТЯ

ДОМОГОСПОДАРСТВ НА ГЕНЕРАЛЬНУ СУКУПНІСТЬ

Формування системи статистичних ваг для науково – обґрунтованого поширення результатів вибіркового обстеження на генеральну сукупність є одним з основних етапів обробки результатів ОУЖД. Важливість цього етапу обумовлена тим, що некоректна система ваг знижує якість оцінок показників для генеральної сукупності, розрахованих за даними вибіркового обстеження. Статистичне зважування забезпечує можливість врахування загальних ймовірностей включення домогосподарств до вибірки, підвищує рівень репрезентативності вибірки в плані відображення нею кількісних та структурних характеристик генеральної сукупності та дозволяє ефективно вирішувати такі проблеми, як врахування нерівних ймовірностей відбору одиниць вибірки, фактичних рівнів відмов домогосподарств від участі в обстеженні по окремих територіальних одиницях або групах домогосподарств, узгодження результатів обстеження з достовірною зовнішньою інформацією. Таким чином, система статистичних ваг відображає дизайн вибірки, а також враховує основні етапи реалізації обстежень.

Дизайн вибірки обстеження умов життя домогосподарств в Україні розроблено у відповідності з сучасною міжнародною практикою проведення обстежень доходів і витрат домогосподарств [2, 3]. Формування вибірки здійснено на основі процедури стратифікованого багатоступеневого ймовірнісного (випадкового) відбору. Процедура формування вибірки складається з таких основних етапів:

- 1) виключення територій, що не можуть бути обстежені;
- 2) виключення населення, яке не підлягає обстеженню;
- 3) стратифікація генеральної сукупності;

4) відбір територіальних одиниць;

5) відбір домогосподарств.

У міській місцевості процедура формування вибірки складається з трьох, у сільській – з двох ступенів.

В даній методиці наведений загальний опис методології формування вибірки ОУЖД для циклу 2004 – 2008 рр. Методологічні засади та детальний опис процедур та алгоритмів формування вибірки для ОУЖД наведено в [4, 5].

При формуванні територіальної вибірки були виключені населені пункти, розташовані у зоні відчуження (I зона) та зоні безумовного (обов'язкового) відселення (II зона) на території, що зазнала радіоактивного забруднення внаслідок Чорнобильської катастрофи. Відповідно, чисельність населення, яка проживає на цих територіях, також була виключена із чисельності населення України і відповідних областей.

При здійсненні розрахунків чисельності населення, яке підлягає обстеженню, з чисельності постійного населення виключено інституційне населення – військовослужбовці строкової служби; особи, які знаходяться у місцях позбавлення волі; особи, які постійно проживають у будинках-інтернатах, будинках для осіб похилого віку тощо. Крім того, вибіркова сукупність не репрезентує маргінальні прошарки населення (безпритульні тощо).

Стратифікація генеральної сукупності здійснювалася з метою адекватного відображення у вибірці основних особливостей адміністративно-територіального поділу України, а також для забезпечення відбору з більш однорідних за основними характеристиками сукупностей домогосподарств [4]. Для реалізації цієї мети генеральна сукупність була розподілена на страти, які відповідають міським поселенням (міськрадам) з чисельністю населення 100 тис. осіб і більше (далі великі міста), міським поселенням (міськрадам та селищним радам) з чисельністю населення менше 100 тис. осіб (далі малі міста) та районам (крім внутрішньоміських), тобто адміністративним районам сільської місцевості (далі райони) за регіонами України. Таким чином, по Україні сформовано 52 страти по міських поселеннях і 490 страт по сільській місцевості.

Обсяг вибірки розподілявся між стратами пропорційно чисельності неінституційного постійного населення. При цьому по міських поселеннях враховувалось все населення (міське та сільське), що проживає в межах поселень, по районах враховувалось лише сільське населення.

При формуванні вибіркової сукупності домогосподарств для ОУЖД у 2004-2008 роках на першому ступені відбору з ймовірністю, пропорційною розміру, відбиралися первинні територіальні одиниці вибірки (ПТОВ): в межах страт по міських поселеннях – великі та малі міста, в межах страт по сільській місцевості – сільські ради. На другому ступені відбору у міських ПТОВ з ймовірністю, пропорційною розміру, відбиралися вторинні територіальні одиниці вибірки (ВТОВ) – інструкторські дільниці Всеукраїнського перепису населення 2001 року, у сільських ПТОВ з використанням процедури систематичного відбору відбиралися адреси домогосподарств. З використанням цієї ж процедури адреси домогосподарств по міських ВТОВ відбиралися на третьому ступені відбору.

В імовірнісній вибірці кожне j -те домогосподарство генеральної сукупності обсягу N , $j = 1, 2, \dots, N$, повинне мати певну позитивну ймовірність p_j потрапити до вибірки. Інакше кажучи, за результатами відбору це домогосподарство буде репрезентувати $1/p_j$ одиниць генеральної сукупності і, відповідно, при оцінюванні показників по генеральній сукупності його потрібно врахувати $1/p_j$ разів. Таким чином, статистична вага одиниці вибірки при поширенні результатів обстеження на генеральну сукупність розраховується за наступною формулою:

$$w_{Bj} = 1/p_j. \quad (1.1)$$

Ймовірність p_j є загальною ймовірністю включення домогосподарств до вибірки, а вага w_{Bj} , розрахована на її основі, є базовою вагою. Базова вага визначається дизайном вибірки, точніше ймовірностями формування вибірок даного дизайну та обсягу [2, 3, 6].

У кінцевому масиві первинних даних, отриманих за результатами вибіркового обстеження, як правило, представлені дані не по відібраних одиницях, а по одиницях, за якими вдалося отримати інформацію. Причинами цього є такі явища, як тривала відсутність членів домогосподарств вдома, відмови від інтерв'ю та ін. Таким чином, кінцева система ваг повинна відображати не лише ймовірності включення одиниць до вибірки, а ще й ймовірність отримання інформації по кожній з відібраних одиниць.

У загальному випадку кінцева вага i -го домогосподарства вибірки обсягу n , $i = 1, 2, \dots, n$, розраховується як добуток базової ваги цієї одиниці та відповідних коефіцієнтів коригування ваг (коефіцієнтів переважування) за формулою:

$$w_i = w_{Bi} \cdot k_{1i} \cdot k_{2i} \cdot \dots \cdot k_{Ei}, \quad (1.2)$$

де w_i – кінцева вага i -ї одиниці; w_{Bi} – базова вага i -ї одиниці; $k_{1i}, k_{2i}, \dots, k_{Ei}$ – коефіцієнти коригування ваг; E – загальна кількість етапів коригування системи ваг.

Система коефіцієнтів коригування ваг повинна враховувати ефекти впливу різних факторів на кінцеві ваги домогосподарств.

Таким чином, кінцеві статистичні ваги – це позитивні числа, розраховані за спеціальною процедурою для кожного фактично обстеженого домогосподарства, що характеризують співвідношення певних характеристик частини генеральної сукупності, яку репрезентує це домогосподарство, і характеристик самого домогосподарства або групи домогосподарств, до якого воно відноситься.

Для ОУЖД 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 одиниць;
- коригування ваг для врахування ступеня участі відібраних одиниць в обстеженні;

- коригування ваг для узгодження результатів обстеження з наявною зовнішньою інформацією та з метою забезпечення однаковості ваг осіб, що є членами одного домогосподарства;
- контроль якості системи ваг.

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

Базова вага, що визначається за формулою (1.1), – це результат певної послідовності розрахунків. Існуючий дизайн вибірки ОУЖД обумовлює використання для побудови базових ваг наступних складових:

- імовірність відбору ПТОВ P_{1i} ;

- імовірність відбору одиниць другого ступеню P_{2i} – ВТОВ по міських поселеннях та домогосподарств по сільській місцевості;

- імовірність відбору домогосподарств у межах ВТОВ P_{3i} по міських поселеннях.

Базовою вагою \bar{i} -го домогосподарства з вибірки є обернена величина добутку зазначених ймовірностей:

$$w_{Bi} = \frac{1}{P_i} = \begin{cases} \frac{1}{P_{1i} \cdot P_{2i} \cdot P_{3i}}, & \text{по міських поселеннях,} \\ \frac{1}{P_{1i} \cdot P_{2i}}, & \text{по сільській місцевості,} \end{cases} \quad (1.3)$$

Слід зазначити, що величина загальної імовірності включення одиниць до вибірки P_i визначалася виходячи з того, що кожне домогосподарство при відборі повинно мати однакову ймовірність потрапити до вибірки. Вибірка, яка побудована з урахуванням зазначеної умови, називається самозваженою (або рівнозваженою). Ця умова відображається наступною формулою:

$$P_i = f = \begin{cases} P_{1i} \cdot P_{2i} \cdot P_{3i}, & \text{по міських поселеннях,} \\ P_{1i} \cdot P_{2i}, & \text{по сільській місцевості,} \end{cases} \quad (1.4)$$

де f – частка відбору, що дорівнює відношенню обсягу вибірки до загальної кількості домогосподарств у

генеральній сукупності, $f = \frac{n}{N}$. Величина, обернена до f , показує скільки домогосподарств генеральної сукупності репрезентує одне відібране домогосподарство при плануванні вибірки.

Для забезпечення самозваженості вибірки визначається теоретична кількість домогосподарств $n_{K(i)}$, що повинна відбиратися на останньому етапі формування вибірки по територіальній одиниці h , до якої належить \bar{i} -те домогосподарство.

$$P_{3i} = \frac{n_{K(i)}}{N_{2i}}$$

У відповідності з формулою (1.4), враховуючи, що по міських поселеннях

$$P_{2i} = \frac{n_{K(i)}}{N_{1i}}$$

місцевості N_{1i} , кількість домогосподарств, що відбирається на останньому етапі формування

вибірки $n_{K(i)}$, визначається за наступним теоретичним співвідношенням:

$$n_{K(i)} = \begin{cases} \frac{P_i \cdot N_{2i}}{P_{1i} \cdot P_{2i}} = \frac{f \cdot N_{2i}}{P_{1i} \cdot P_{2i}}, & \text{по міських поселеннях,} \\ \frac{P_i \cdot N_{1i}}{P_{1i}} = \frac{f \cdot N_{1i}}{P_{1i}}, & \text{по сільській місцевості,} \end{cases} \quad (1.5)$$

де N_{2i} – загальна кількість домогосподарств у територіальній одиниці другого ступеню, до якої належить \bar{i} -те домогосподарство, по міських поселеннях; N_{1i} – загальна кількість домогосподарств у територіальній одиниці першого ступеню, до якої належить \bar{i} -те домогосподарство, по сільській місцевості.

Базові ваги розраховуються на весь термін експлуатації вибірки. Територіальна вибірка для ОУЖД розрахована для експлуатації протягом п'яти років, тобто протягом цього періоду не змінюються її дизайн і обсяг.

Теоретична кількість одиниць, яку необхідно відібрати при побудові вибірки, і яка забезпечує самозваженість вибірки, часто залишається саме теоретичною, оскільки на практиці доводиться враховувати такі фактори, як умови роботи працівників, які проводять опитування домогосподарств за програмою обстеження (надалі – статистики з обстежень), наприклад, забезпечувати рівномірну їх завантаженість, а також виправляти певні похибки в основі вибірки або у процедурах відбору. Тому необхідно скоригувати базові ваги з урахуванням

фактичної кількості відібраних одиниць. Це здійснюється шляхом введення спеціального коефіцієнту врахування фактичної кількості відібраних одиниць.

При побудові вибірки доцільно намагатися, щоб коефіцієнт врахування фактичної ймовірності відбору одиниць був близьким до одиниці. В іншому випадку вибірка не буде самозваженою, що призведе до зниження статистичної ефективності обстеження.

Для підвищення якості результатів обстежень система ваг має забезпечити врахування відхилень від теоретичного дизайну вибірки, які виникають під час реалізації обстеження і мають об'єктивний характер, зокрема, це наявність зачинених та нежитлових приміщень, відмови окремих домогосподарств від участі в обстеженні.

Одним з найбільш розповсюджених і теоретично обґрунтованих підходів до врахування випадків неодержання відповідей при обробці даних вибіркового обстеження є коригування статистичних ваг, сутність якого полягає в завищенні ваг одиниць, які мають характеристики, аналогічні характеристикам одиниць, що не були обстежені.

Коригування системи ваг для компенсації відмов домогосподарств від обстеження має суттєве значення в дослідженнях, яким притаманні високі ступені відмов та їх різний рівень серед окремих груп одиниць у вибірці. Це призводить до необхідності коригування системи ваг для запобігання суттєвим зміщенням у розподілі часток проведених інтерв'ю по групах домогосподарств вибірки.

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 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 зовнішньою інформацією щодо певних характеристик генеральної сукупності здійснюється спеціальне коригування системи статистичних ваг. Доцільність цього обумовлена наступним:

- 1) оцінки показників за результатами вибіркового обстеження завжди характеризуються певною похибкою, джерелами якої є несутільність обстеження (похибка вибірки);
- 2) вибіркоче 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. Коригування статистичних ваг здійснюються з метою наближення оцінок окремих показників, отриманих у вибіркового обстеженні, до актуальних надійних зовнішніх даних;
- 3) основні характеристики генеральної сукупності змінюються у часі, і, відповідно, окремі з них на момент завершення обробки даних поточного циклу обстеження дещо відрізняються від таких, якими вони були під час початку п'ятирічного циклу дії територіальної вибірки в обстеженні.

Крім перелічених проблем, часто виникає ситуація, коли необхідно узгоджувати результати обстеження для двох рівнів одиниць спостереження – для домогосподарств і осіб. Насамперед – це узгодження окремих кількісних характеристик, що можуть бути розраховані як по домогосподарствах, так і по особах. Наприклад, бажано, щоб чисельність населення, оцінена для рівня домогосподарств (за їх розміром), дорівнювала чисельності, оціненою для рівня осіб.

З відомих в теперішній час узагальнених методів коригування системи статистичних ваг найбільш теоретично розробленою та ефективною є так звана процедура калібрації статистичних ваг [6 - 8].

Термін "калібрація" означає трансформацію (коригування) системи ваг одиниць вибірки шляхом мінімізації функції відстані між калібраційними вагами та вагами, що коригуються, за умови, що калібраційні ваги дають точні оцінки зовнішніх даних, відомих для генеральної сукупності або її частин.

З математичної точки зору калібраційний метод полягає в тому, що розв'язується задача мінімізації з обмеженнями, де обмеження – це умови відповідності оцінок показників наявній зовнішній інформації. Дещо спрощено постановка цієї задачі може бути представлена у наступному вигляді:

$$\sum_{i=1}^n \frac{1}{2} \frac{(w_i^{(d)} - w_i^{(c)})^2}{w_i^{(c)}} \rightarrow \min, \quad (1.6)$$

при дотриманні обмеження:

$$\sum_{i=1}^n w_i^{(c)} x_i = X \quad (1.7)$$

де $w_i^{(d)}$ – ваги, що калібруються (ваги дизайну); $w_i^{(c)}$ – калібраційні ваги (ваги після калібрації); x_i – змінна, що вимірюється в обстеженні і, за якою наявна зовнішня достовірна інформація. В даному випадку приймається, що з додаткових джерел відоме сумарне значення X . Прикладом ознаки x_i може бути розмір домогосподарства. Тоді X – це загальна чисельність населення.

Таким чином, мінімізується відстань між вагами, що коригуються, $w_i^{(d)}$ та калібраційними вагами $w_i^{(c)}$ за умовою дотримання певних обмежень.

Розв'язання задачі (1.6) – (1.7) може бути записано у вигляді:

$$w_i^{(c)} = w_i^{(d)} \cdot \left(1 + \left(X - \sum_{i=1}^n w_i^{(d)} x_i \right) \cdot \left(\sum_{i=1}^n w_i^{(d)} x_i^2 \right)^{-1} x_i \right) = w_i^{(d)} \cdot k_i^{(c)} \quad (1.8)$$

де $k_i^{(c)}$

– коефіцієнт коригування ваг за результатами калібрації.

Слід зазначити, що застосування процедури калібрації має і певні недоліки. Основні з них такі:

– може погіршуватись якість системи ваг, зокрема, для окремих домогосподарств з вибірки можуть продукуватись нульові або від'ємні ваги;

– може погіршуватись якість оцінок окремих показників.

Крім того, застосування методу калібрації вимагає вирішення проблем з визначенням зовнішньої інформації належної якості та з вибором оптимальної кількості зовнішніх джерел даних.

Процедура калібрації статистичних ваг, як математичний алгоритм, може працювати з будь-якою системою ваг і майже з будь-яким набором зовнішніх даних. Але для вирішення проблеми підвищення якості результатів обстеження, зокрема шляхом забезпечення їх узгодженості з зовнішніми даними, необхідно мати певний досвід з питань застосування процедур коригування системи ваг, визначення зовнішніх джерел даних та якості інформації, яку вони містять. Таким чином, калібрація ваг не є формальною процедурою. Застосування цих методів в практичній діяльності вимагає виконання спеціальних досліджень та високої кваліфікації працівників.

В теперішній час в ОУЖД калібрація статистичних ваг здійснюється на основі рекомендацій міжнародних консультантів з використанням спеціально розроблених та стандартних програмних засобів.

Враховуючи вплив системи ваг на якість оцінок показників, отриманих за результатами ОУЖД, постає проблема жорсткого контролю за якістю самих 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 відхилена від стану самозваженості за рахунок таких факторів, як наявність у списках для відбору адрес неіснуючих домогосподарств, відхилення навантаження статистиків з проведення вибіркових обстежень населення, від теоретично визначеної величини, відмови домогосподарств від участі в обстеженні, що мають різний рівень в різних територіальних одиницях тощо. В окремих випадках, ці ефекти можуть мати настільки значний вплив на систему ваг, що необхідно відмовитися від принципу самозваженості та диференціювати частки відбору одиниць по територіальних одиницях з метою зменшення ступеня варіації кінцевих ваг і, відповідно, поліпшення оцінювання показників.

Таким чином, у реальних обстеженнях необхідно аналізувати якість системи ваг і використовувати всі можливості для її поліпшення.

2. ПОРЯДОК РОЗРАХУНКУ СКЛАДОВИХ СИСТЕМИ ВАГ

2.1. Розрахунок базової ваги домогосподарств

2.1.1. Розрахунок базової ваги домогосподарств міських поселень

Визначення імовірності відбору одиниць першого ступеня P_{li}

А. Саморепрезентативні міські поселення

Великі міста

Саморепрезентативними є одиниці відбору першого ступеня з розміром (чисельністю населення) вище

встановленого порогу саморепрезентативності. Для таких одиниць відбору $P_{li} = 1$. Порогове значення

чисельності n_p визначається за формулою:

$$n_p = \frac{N}{n} \cdot m \cdot m_f \quad (2.1)$$

де n – обсяг вибірки, домогосподарств; m – навантаження статистика (кількість домогосподарств, яка має бути обстежена одним статистиком з проведення вибіркових обстежень населення протягом одного кварталу),

домогосподарств; m_f – середній розмір домогосподарства, осіб; N – обсяг генеральної сукупності, домогосподарств.

При формуванні вибірки ОУЖД розрахунок порогу саморепрезентативності здійснювався на основі параметрів державного обстеження робочої сили (ОРС), враховуючи, що територіальні одиниці цих обстежень по міських поселеннях співпадають, і навантаження статистика в ОУЖД є меншим (для великих міст: 72 домогосподарства в ОРС проти 28 в ОУЖД). Для міських поселень з чисельністю населення 100 тис. осіб параметри

ОРС мають наступні значення: загальна кількість неінституційних домогосподарств в Україні $N = 17609000$; обсяг вибірки ОРС $n_{ОРС} = 31800$; навантаження статистика в ОРС $m_{ОРС} = 72$ домогосподарства; середній розмір

домогосподарства $m_f = 2,62$ особи. Поріг саморепрезентативності:

$$n_p = 553,75 \cdot 72 \cdot 2,62 \approx 104500 \text{ осіб} \quad (2.2)$$

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

Малі міста

Враховуючи, що за принципами організації ОУЖД та ОРС навантаження статистиків у малих містах нижче, ніж у великих, для ОРС відповідно 54 і 72 домогосподарства, то необхідно виділити саморепрезентативні малі міста, які також включаються до вибірки з ймовірністю, рівною одиниці. Необхідно зазначити, що для таких малих міст, як і

для великих міст, при визначенні порогу самореферентивності були використані параметри ОРС, беручи до уваги що навантаження статистика для малих міст в ОРС (54 домогосподарства) вище ніж в ОУЖД (21 домогосподарство).

Для малих міст $m_{ОРС}=54$ домогосподарства; середній розмір домогосподарства $m_f = 2,65$ особи

Поріг самореферентивності n_{p2} :

$$n_{p2} = 55375 \cdot 54 \cdot 2,65 \approx 79242 \text{ особи} \quad (2.3)$$

Таким чином, міські поселення з чисельністю населення $n \geq 79,2$ тис. осіб. (самореферентивні малі міста) включалися у вибірку з ймовірністю, рівною одиниці.

Б. Несамореферентивні міські поселення
Несамореферентивні міські поселення відбирались по кожній страті окремо. Кількість несамореферентивних міст a , яку необхідно відібрати у регіоні, визначалась шляхом ділення чисельності населення всіх несамореферентивних міст регіону на поріг самореферентивності n_{p2} .

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

$$P_{li} = \left(a \cdot \frac{M_a}{\sum M_a} \right), \quad (2.4)$$

де: a – кількість відібраних малих міст у регіоні; M_a – чисельність населення міста, для якого

розраховується P_{li} ; $\sum M_a$ – загальна чисельність населення несамореферентивних малих міст у регіоні.

Результати розрахунку імовірностей включення до вибірки несамореферентивних малих міст розглянемо на прикладі Автономної Республіки Крим. За результатами відбору до вибірки включені сім міських поселень: Алушта, Джанкой, Саки, Бахчисарай, Нижньогірський, Зуя, Поштове (див. табл. 2.1., рис. 2.1). Чисельність населення всіх несамореферентивних малих міст АР Крим складає 446300 осіб. Крок відбору міських поселень (фактичне порогове значення для АР Крим) складає 63757 осіб (= 446300/7). Імовірності відбору міських поселень наведені у табл. 2.1.

Різниця між кроком відбору 63757 осіб і порогом самореферентивності 79182 осіб виникла за рахунок округлення кількості міських поселень, які необхідно відібрати, до цілого значення. Потенційно таке округлення може бути джерелом певного зростання дисперсії кінцевої системи ваг.

Величина P_{li} для кожного відбраного міста характеризує частку відрізка загальною довжиною 63757 осіб, яка відповідає чисельності населення цього міста (рис. 2.1).

Таблиця 2.1.

Результати відбору несамореферентивних міських поселень в Автономній Республіці Крим для ОРС та ОУЖД у 2004-2008 роках

| Міське поселення | Чисельність населення, осіб | P_{li} |
|------------------|-----------------------------|----------|
| Алушта | 53000 | 0,8313 |
| Джанкой | 42400 | 0,6650 |
| Саки | 28500 | 0,4470 |
| Бахчисарай | 27100 | 0,4251 |
| Нижньогірський | 11900 | 0,1866 |
| Зуя | 10100 | 0,1584 |
| Поштове | 8300 | 0,1302 |

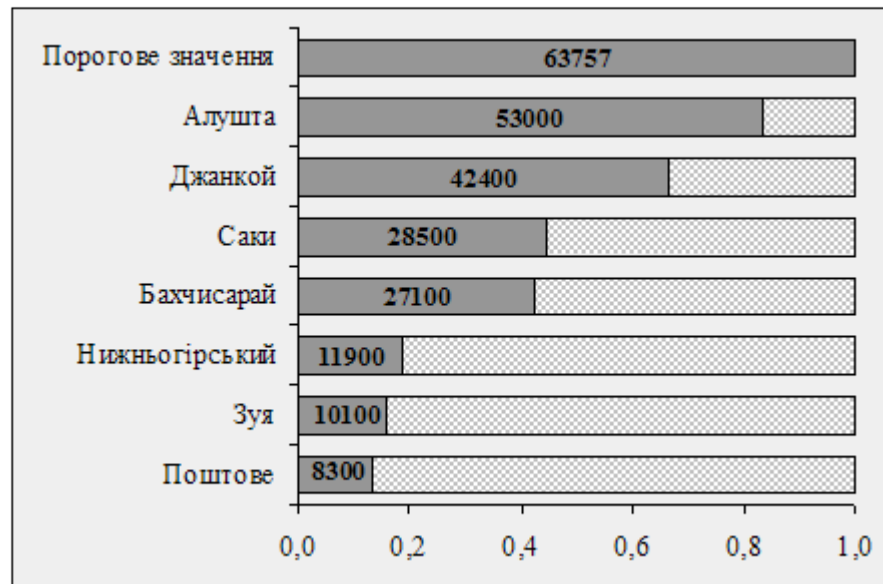


Рис. 2.1. Імовірності включення до вибірки несамопрезентативних міських поселень Автономної Республіки Крим для ОРС та ОУЖД у 2004 – 2008 роках

Якщо розрахувати вагу домогосподарств (населення) певного міста q в межах регіону, то вона дорівнюватиме $W(q) = 1/P_1(q)$. Для прикладу, що розглядається, зміст цієї величини буде полягати в наступному: це кількість домогосподарств (осіб) страти, яку представляє одне домогосподарство (одна особа) відібраного міста q .

Визначення імовірності відбору одиниць другого ступеня P_{2i}

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

$$P_{2i} = \left(b \cdot \frac{M_B}{\sum_B M_B} \right), \quad (2.5)$$

де: b – кількість ВТОВ, відібраних у конкретному місті; M_B – чисельність постійного населення за даними

перепису у ВТОВ, для якої розраховується P_{2i} ; $\sum_B M_B$ – загальна чисельність населення у відібраному 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 самопрезентативна.

Загальна ймовірність включення до вибірки ВТОВ розраховується як добуток ймовірностей P_{1i} і P_{2i} для кожної ВТОВ. Загальні ймовірності відбору помітно різняться для різних ВТОВ (див. рис. 2.2). При формуванні вибірки необхідно аналізувати загальні ймовірності відбору ВТОВ для виявлення можливих похибок, які відображаються, зокрема у вигляді окремих нестандартних значень (аутлайерів), які суттєво відрізняються від середнього значення ймовірності і можуть призвести до погіршення якості системи ваг.

У випадку виявлення вищезгаданих аутлайерів застосовуються спеціальні процедури їх коригування, зокрема процедури імпутації загальних ймовірностей відбору ВТОВ, які використовуються замість нестандартних значень.

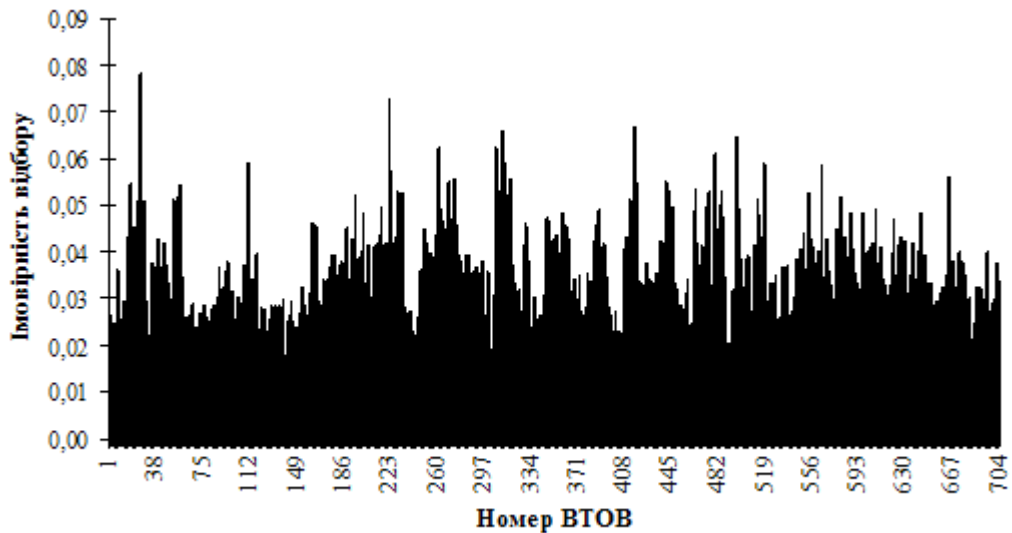


Рис. 2.2 Загальні імовірності відбору ВТОВ по міських поселеннях для ОРС та ОУЖД у 2004-2008 роках

Визначення імовірності відбору домогосподарств P_{3i}

Теоретична імовірність відбору i -го домогосподарства в межах відібраної ВТОВ розраховується виходячи з формули (див. (1.4)):

$$P_{3i} = \frac{f}{(P_{1i} \cdot P_{2i})} \quad (2.6)$$

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

$$w_{Bi} = \frac{1}{P_{1i} \cdot P_{2i} \cdot P_{3i}} = \frac{1}{f} \quad (2.7)$$

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:

$$P_{1i} = \left(c \cdot \frac{M_c}{\sum_c M_c} \right) \quad (2.8)$$

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. (1.4.)) :

$$P_{2i} = \frac{f}{P_{1i}} \quad (2.9)$$

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

$$w_{Bi} = \frac{1}{P_{1i} \cdot P_{2i}} = \frac{1}{f} \quad (2.10)$$

2.1.3. Розрахунок частки відбору домогосподарств f

При формуванні вибіркової сукупності домогосподарств для ОУЖД у 2004 році частка відбору

$$f = \frac{n_{\text{ОУЖД}}}{N} = \frac{12977}{17609200}$$

дорівнювала . Ця величина розрахована як відношення обсягу вибіркової сукупності ОУЖД (визначеного з урахуванням рівня надійності результатів обстеження, який необхідно забезпечити, вартості обстеження та очікуваного рівня участі домогосподарств 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 ОУЖД та ін.). Таким чином, базова

$$w_{Bi} = \frac{1}{f} \approx 1357$$

вага домогосподарств в ОУЖД дорівнює

Доцільно зазначити наступне. Величина частки відбору домогосподарств f за своїм змістом суттєво відрізняється від величини загальної ймовірності включення домогосподарства до вибірки P_i , хоча для самозваженої вибірки ОУЖД і приймається, що $P_i = f$. Величина f використовується для загальної характеристики вибірки, її

дизайну. При розрахунку системи статистичних ваг використовується насамперед величина P_i та її складові. Це пояснюється тим, що, в процесі побудови вибірки необхідно контролювати величини ймовірностей відбору одиниць різних ступенів, оцінювати відхилення теоретичних ймовірностей відбору домогосподарств від фактичних тощо. Крім того, при формуванні інших вибірок (або підвибірок) домогосподарств або осіб на основі територіальної вибірки ОУЖД, може змінитися, наприклад, принцип визначення ймовірностей відбору домогосподарств, що призведе до необхідності використання ймовірностей відбору ПТОВ і ВТОВ для розрахунку відповідної нової загальної ймовірності включення домогосподарств до вибірки. При необхідності розрахунку показників ОУЖД, наприклад, по окремих несамопрезентативних містах, загальна ймовірність включення домогосподарства до вибірки ОУЖД повинна бути розрахована без урахування ймовірності відбору ПТОВ.

2.2. Врахування фактичної ймовірності відбору домогосподарств

Фактичні ймовірності відбору домогосподарств в межах територіальних одиниць можуть відрізнятися від теоретичних, оскільки під час реалізації обстежень, як вже зазначалося у п. 1, необхідно враховувати умови роботи статистиків, зокрема забезпечувати рівномірну їх завантаженість. Наприклад, в обстеженні ОУЖД прийнято, що у великих містах навантаження статистика з проведення вибірових обстежень населення складає 28 домогосподарств, він обстежує дві ВТОВ, в кожній з яких відбирається по 14 домогосподарств, а у малих містах – 21 домогосподарство, статистик обстежує дві ВТОВ, в одній з яких відбирається 11 домогосподарств, а в іншій – 10 (див. рис.2.3).

Після відбору домогосподарств необхідно скоригувати базові ваги для врахування фактичної ймовірності відбору домогосподарств. Таке коригування здійснюється з використанням відповідного коефіцієнту коригування

ваг k_{1i} , який розраховується за формулою:

$$k_{1i} = \begin{cases} \frac{P_i}{P_f} & \text{по міських поселеннях} \\ \frac{P_i}{P_f} & \text{по сільській місцевості} \end{cases} \quad (2.11)$$

де P_f – фактична ймовірність відбору домогосподарства в межах територіальної одиниці.

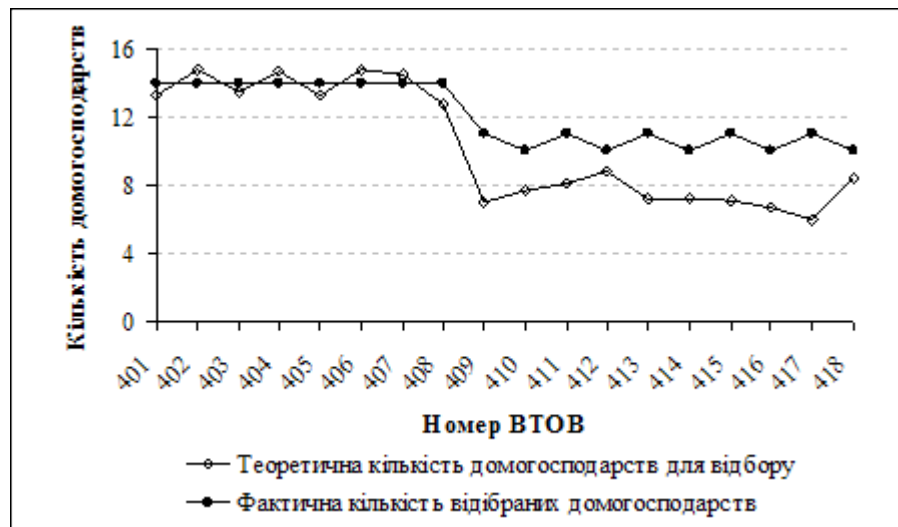


Рис. 2.3. Фактична кількість відібраних домогосподарств та теоретична кількість домогосподарств для відбору по міських ВТОВ Миколаївської області для ОУЖД у 2004-2008 рр.

Фактичні ймовірності відбору домогосподарств в межах територіальних одиниць P_{fi} розраховуються за формулою:

$$P_{fi} = \frac{n_{TO}}{N_{TO}}, \quad (2.12)$$

де n_{TO} – фактична кількість домогосподарств, що відібрані для обстеження у територіальній одиниці (ВТОВ по міських поселеннях або ПТОВ по сільській місцевості); N_{TO} – загальна кількість домогосподарств у територіальній одиниці.

Після визначення коефіцієнтів коригування ваг k_{1i} розраховуються нові ваги домогосподарств w_{1i} за формулою:

$$w_{1i} = w_{fi} \cdot k_{1i}. \quad (2.13)$$

Статистичні ваги після врахування фактичних ймовірностей 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 (класів зважування). 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.

Для ОУЖД комірки формуються за такими класифікаційними змінними:

- економічний район – 8 дискретних значень (економічний район - група регіонів, що об'єднані за близькістю таких характеристик, як виробництво промислової продукції, концентрація виробничого потенціалу та робочої сили [9]);
- тип місцевості – 3 дискретних значення (велике місто, мале місто, село);
- тип домогосподарства за розміром – 4 дискретних значення (одинаки, домогосподарства, що складаються з двох осіб, з трьох осіб, з чотирьох осіб і більше).

Важливо зазначити, що інформацію за наведеними класифікаційними змінними статистик з проведення вибіркового обстеження населення збирає з різних джерел, в т.ч. і по домогосподарствах, які відмовилися від обстеження.

Код комірки розраховується за формулою:

$$c_i = \text{економічний район} \cdot 100 + \text{тип місцевості} \cdot 10 + \text{тип домогосподарства за розміром} \quad (2.14)$$

Для побудови комірок при розрахунку системи ваг за результатами основного інтерв'ю по обстежених домогосподарствах використовуються дані опитування респондентів щодо розміру домогосподарства, а для домогосподарств, які відмовилися від обстеження, застосовуються дані, отримані статистиком з додаткових джерел.

Комірки нумеруються починаючи з одиниці після ранжування їх за кодом.

Загальна початкова кількість домогосподарств у i -ій комірці n_i при обробці даних основного інтерв'ю визначається за результатами проведення відбору домогосподарств. При цьому загальною кількістю домогосподарств у комірці виключаються нежитлові приміщення, зачинені приміщення та неіснуючі адреси.

За результатами проведення кожного квартального інтерв'ю будується своя система комірок. Загальна

початкова кількість домогосподарств у i -ій комірці при обробці даних квартальних інтерв'ю визначається за результатами проведення основного інтерв'ю та попередніх квартальних інтерв'ю поточного року ОУЖД.

Кількість відмов домогосподарств від обстеження по комірках визначається за результатами проведення основного та квартальних інтерв'ю ОУЖД.

Якщо у деякій комірці за результатами опитування кількість обстежених домогосподарств менша, ніж 20 (дане значення відповідає наявним експертним оцінкам, але не є загальноприйнятим стандартом, якого в теперішній час не існує), ця комірка об'єднується з суміжною за типом домогосподарства за розміром. При цьому економічний район та тип місцевості у комірок, що об'єднуються, повинні бути ідентичні.

Коефіцієнт врахування відмов домогосподарств від інтерв'ю k_{2i} розраховується по кожній i -ій комірці за результатами основного та кожного квартального інтерв'ю ОУЖД за формулою:

$$k_{2i} = \frac{\sum_{f=1}^{\pi} w_{1i} \cdot \lambda_f}{\sum_{f=1}^{\pi} w_{1i} \cdot \lambda_f \cdot \eta_f}, \quad (2.15)$$

$$\lambda_f = \begin{cases} 1, & \text{якщо } f \in M_i; \\ 0, & \text{якщо } f \notin M_i; \end{cases}$$

де

$$\eta_i = \begin{cases} 1, & \text{якщо домогосподарство взяло участь в обстеженні} \\ 0, & \text{якщо домогосподарство відмовило від обстеження} \end{cases}$$

M_i – множина домогосподарств i -ої комірки для основного або певного квартального інтерв'ю ОУЖД.

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

Після визначення коефіцієнтів коригування ваг k_{2i} розраховуються ваги домогосподарств з урахування відмов від обстеження w_{2i} за формулою:

$$w_{2i} = w'_{1i} \cdot k_{2i}. \quad (2.16)$$

Ваги w_{2i} визначаються за результатами обробки даних основного або окремо кожного квартального інтерв'ю ОУЖД. При цьому для даних основного інтерв'ю w'_{1i} – це ваги з урахуванням фактичної імовірності відбору домогосподарств w_{1i} , для даних першого кварталу – кінцеві ваги основного інтерв'ю, а для даних другого, третього та четвертого кварталів – це кінцеві ваги відповідно першого, другого та третього кварталів.

2.4. Калібрація системи статистичних ваг

Для узгодження результатів, що отримані за даними основного інтерв'ю та за даними кожного з чотирьох квартальних інтерв'ю річного циклу ОУЖД, з наявною зовнішньою інформацією застосовується процедура калібрації системи статистичних ваг.

Зовнішні дані для калібрації визначаються на основі методичних підходів, які частково висвітлені в [6], і передбачають проведення спеціальних досліджень. Для циклу обстежень ОУЖД у 2004 – 2008 роках були визначені наступні основні дані, які отримані з зовнішніх джерел і використовуються як оцінки, що необхідно забезпечити за результатами обстежень:

- загальна кількість домогосподарств за регіонами та типами місцевості;
- кількість домогосподарств з дітьми за регіонами;
- чисельність неінституційного населення за регіонами та типами місцевості;
- структура населення за регіонами та статеві-віковими групами.

У додатку А наведена зовнішня інформація, що використовувалась в ОУЖД для калібрації системи ваг у 2005 році. Звернемо увагу, що зовнішня інформація може бути визначена для різних рівнів агрегації даних – для національного або регіонального рівнів та за різними розрізами.

Постановка задачі калібрації для ОУЖД у 2005 році, що відповідає наведеній у додатку А структури зовнішніх даних, в межах окремого регіону має вигляд:

$$\left\{ \begin{array}{l} \sum_{i=1}^n \frac{1}{2} \frac{(w_{3i} - w_{2i})^2}{w_{2i}} \rightarrow \min; \\ \sum_{i=1}^n j_i^{(b-t)} w_{3i} = H^{(b-t)}; \\ \sum_{i=1}^n j_i^{(s-t)} w_{3i} = H^{(s-t)}; \\ \sum_{i=1}^n j_i^{(rw)} w_{3i} = H^{(rw)}; \\ \sum_{i=1}^n h_i^{(b-t)} w_{3i} = N^{(b-t)}; \\ \sum_{i=1}^n h_i^{(s-t)} w_{3i} = N^{(s-t)}; \\ \sum_{i=1}^n h_i^{(rw)} w_{3i} = N^{(rw)}; \\ \sum_{i=1}^n d_i w_{3i} = D; \\ \sum_{i=1}^n g_i^{(1)} w_{3i} = G^{(1)}; \\ \sum_{i=1}^n g_i^{(2)} w_{3i} = G^{(2)}; \\ \sum_{i=1}^n g_i^{(3)} w_{3i} = G^{(3)}; \\ \sum_{i=1}^n g_i^{(4)} w_{3i} = G^{(4)}; \\ \sum_{i=1}^n m_i^{(1)} w_{3i} = M^{(1)}; \\ \sum_{i=1}^n m_i^{(2)} w_{3i} = M^{(2)}; \\ \sum_{i=1}^n m_i^{(3)} w_{3i} = M^{(3)}; \\ \sum_{i=1}^n m_i^{(4)} w_{3i} = M^{(4)}; \end{array} \right.$$

де $w_{2i}, i = 1, \dots, n$ – система статистичних ваг до калібрації;

$w_{3i}, i = 1, \dots, n$ – система калібраційних ваг (система статистичних ваг після калібрації);

$j_i^{(b-t)}$ – структурна змінна: $j_i^{(b-t)} = 1$ для великих міст і $j_i^{(b-t)} = 0$ для малих міст та сільської місцевості;

$j_i^{(s-t)}$ – структурна змінна: $j_i^{(s-t)} = 1$ для малих міст і $j_i^{(s-t)} = 0$ для великих міст та сільської місцевості;

$j_i^{(rw)}$ – структурна змінна: $j_i^{(rw)} = 0$ для великих та малих міст і $j_i^{(rw)} = 1$ для сільської місцевості;

$h_i^{(b-t)}$ – кількість осіб в і-тому домогосподарстві великих міст (для малих міст та сільської місцевості $h_i^{(b-t)}$

= 0);

$h_i^{(s-t)}$ – кількість осіб в і-тому домогосподарстві малих міст (для великих міст та сільської місцевості $h_i^{(s-t)}$ = 0);
 $h_i^{(rw)}$ – кількість осіб в і-му домогосподарстві сільської місцевості (для великих та малих міст $h_i^{(rw)}$ = 0);
 d_i – ознака наявності в і-му домогосподарстві дітей (1 – домогосподарство з дітьми, 0 – домогосподарство без дітей);
 $g_i^{(1)}$ – кількість жінок віком 0-17 років в і-му домогосподарстві;
 $g_i^{(2)}$ – кількість жінок віком 18-35 років в і-му домогосподарстві;
 $g_i^{(3)}$ – кількість жінок віком 36-54 років в і-му домогосподарстві;
 $g_i^{(4)}$ – кількість жінок віком 55 років і більше в і-му домогосподарстві;
 $m_i^{(1)}$ – кількість чоловіків віком 0-17 років в і-му домогосподарстві;
 $m_i^{(2)}$ – кількість чоловіків віком 18-35 років в і-му домогосподарстві;
 $m_i^{(3)}$ – кількість чоловіків віком 36-59 років в і-му домогосподарстві;
 $m_i^{(4)}$ – кількість чоловіків віком 60 років і більше в і-му домогосподарстві;
 $H^{(s-t)}, H^{(s-t)}, H^{(rw)}$ – кількість домогосподарств за зовнішніми даними у великих, малих містах та сільській місцевості регіону відповідно.
 $N^{(s-t)}, N^{(s-t)}, N^{(rw)}$ – загальна чисельність населення в міських поселеннях та у сільській місцевості регіону;
D – кількість домогосподарств з дітьми за зовнішніми даними у регіоні в цілому;
 $G^{(1)}, G^{(2)}, G^{(3)}, G^{(4)}$ – кількість жінок за зовнішніми даними у регіоні в цілому по кожній віковій групі, що розглядаються;
 $M^{(1)}, M^{(2)}, M^{(3)}, M^{(4)}$ – кількість чоловіків за зовнішніми даними у регіоні в цілому по кожній віковій групі, що розглядаються.

При формуванні зовнішньої інформації були прийняті окремі припущення, а саме:

- 1) Статеві-вікова структура інституційного населення (без військовослужбовців строкової служби) ідентична структурі постійного населення за даними демографічної статистики.
- 2) Чисельність неінституційного населення за статеві-віковими групами розраховувалась з урахуванням чисельності та статево-вікової структури інституційного населення.
- 3) Чисельність чоловіків працездатного віку зменшується з урахуванням кількості військовослужбовців строкової служби, розподіленої пропорційно чисельності населення по міських поселеннях та сільській місцевості.
- 4) Чисельність неінституційного населення міських поселень визначається шляхом зменшення чисельності постійного населення з урахуванням чисельності військовослужбовців строкової служби та чисельності інституційного населення.
- 5) Чисельність постійного населення сільської місцевості зменшується з урахуванням чисельності військовослужбовців строкової служби.

Масив зовнішньої інформації для калібрації формується один раз на рік у період обробки даних основного інтерв'ю ОУЖД поточного року. У зазначений період передбачається експертна оцінка масиву зовнішніх даних, включаючи їх якість, та системи припущень, прийнятих при його формуванні.

Перелік джерел зовнішньої інформації, які доцільно використовувати для калібрації, як правило, суттєво обмежений. На цьому етапі обробки даних зовнішня інформація безпосередньо впливає на якість результатів обстеження, оскільки фактично зовнішні дані використовуються вже як дані обстеження.

Головним мотивом використання зовнішньої інформації при розрахунку системи ваг є покращання якості результатів обстеження. В цьому плані слід враховувати, що за наявності багатьох джерел різноманітних даних їх повне використання не обов'язково призведе до досягнення поставленої мети. Частіше за все, необхідно провести спеціальні роботи з відбору джерел та спеціальної підготовки даних. При відборі джерел зовнішніх даних слід враховувати, що використання більшої кількості джерел призводить до у загальному випадку до зростання величини дисперсії вибіркового оцінок.

Основні принципи, які слід враховувати при формуванні масиву зовнішніх даних наступні [11]:

- зовнішні дані повинні пояснювати варіацію основних змінних, що вимірюються за результатами ОУЖД;
- зовнішні дані повинні пояснювати варіацію ймовірностей обстеження домогосподарств;
- класифікаційні ознаки зовнішніх даних відповідають ознакам групування одиниць спостереження за результатами обстеження.

За таких умов основним джерелом інформації стають дані офіційної статистики та надійна адміністративна інформація.

Ефективне використання інструментарію калібрації системи статистичних ваг потребує наявності у фахівців відповідної кваліфікації.

Сформований масив зовнішніх даних використовується без змін при калібрації результатів основного та всіх кварталних інтерв'ю ОУЖД.

Розв'язання задачі (2.17) здійснюється стандартними математичними методами, один з яких – це метод множників Лагранжа.

Слід зазначити, що застосування процедури калібрації має і певні недоліки. Основні з цих недоліків такі:

- може погіршуватись якість системи ваг. Зокрема, можуть продукуватись дуже малі або дуже великі значення ваг для окремих одиниць спостереження, або навіть нульові та негативні значення ваг;
- може погіршуватись надійність оцінювання окремих показників за результатами обстеження;
- виникають проблеми з визначенням зовнішньої інформації належної якості та з вибором оптимальної кількості зовнішніх джерел даних.

Для запобігання погіршенню якості системи ваг та якості оцінювання показників необхідно застосовувати спеціальні контрольні процедури. У випадку виявлення неприпустимих значень ваг або суттєвого погіршення надійності оцінювання показників доцільно переглянути ступінь деталізації зовнішньої інформації, що використовується для калібрації, або її структуру. Можливим є також застосування певних модифікованих процедур калібрації, наприклад, використання функцій відстані іншого, ніж представленого формулою (1.6) типу, формулювання додаткових обмежень на зміну величини статистичних ваг тощо [6].

В теперішній час розроблені сучасні пакети програм для реалізації процедур калібрації системи статистичних ваг. Одним з таких пакетів є g-Calib, який розробляється офісом статистики Бельгії для застосування в середовищі SPSS [6].

На основі процедури калібрації визначаються вагові коефіцієнти k_{3i} , що забезпечують відповідність результатів обстеження встановленим обмеженням, та мінімізують відхилення ваг до та після калібрації (див. 2.17).

Після визначення коефіцієнтів коригування ваг k_{3i} розраховуються калібраційні ваги домогосподарств w_{3i} за формулою:

$$w_{3i} = w_{2i} \cdot k_{3i} \quad (2.18)$$

Зазначимо, що за результатом застосування окремих процедур калібрації може розраховуватись

безпосередньо система калібраційних ваг w_{3i} . За таких умов коефіцієнти коригування ваг k_{3i} доцільно розрахувати окремо за формулою:

$$k_{3i} = \frac{w_{3i}}{w_{2i}} \quad (2.19)$$

Це необхідно для забезпечення однорідності процедур контролю за системою ваг.

Калібраційні ваги w_{3i} визначаються за результатами обробки даних основного та всіх квартальних інтерв'ю ОУЖД. Ці ваги є кінцевими вагами, і використовуються при оцінюванні показників ОУЖД.

У табл. 2.2. для ілюстрації наведені значення ваг та коефіцієнтів коригування ваг для домогосподарств Голопристанського району Херсонської області за результатами основного інтерв'ю ОУЖД 2005 р.

Таблиця 2.2.

Ваги домогосподарств та коефіцієнти коригування ваг для домогосподарств Голопристанського району Херсонської області за результатами основного інтерв'ю ОУЖД 2005

| Код домогосподарства | w_{1i} | k_{2i} | w_{2i} | k_{3i} | w_{3i} |
|----------------------|----------|----------|----------|----------|----------|
| 650266 | 1345,83 | 1,05 | 1412,69 | 0,58 | 823,60 |
| 650267 | 1345,83 | 1,05 | 1412,69 | 0,61 | 866,54 |
| 650268 | 1345,83 | 1,01 | 1353,81 | 0,61 | 830,81 |
| 650269 | 1345,83 | 1,01 | 1353,81 | 1,01 | 1373,42 |
| 650270 | 1345,83 | 1,02 | 1373,42 | 1,19 | 1633,63 |
| 650271 | 1345,83 | 1,01 | 1353,81 | 0,91 | 1236,09 |
| 650272 | 982,08 | 1,01 | 987,90 | 1,32 | 1299,91 |
| 650273 | 982,08 | 1,02 | 1002,21 | 0,83 | 829,30 |
| 650274 | 982,08 | 1,02 | 1002,21 | 1,19 | 1191,77 |
| 650275 | 982,08 | 1,01 | 987,90 | 0,53 | 527,28 |
| 650276 | 982,08 | 1,01 | 987,90 | 1,03 | 1020,93 |
| 650277 | 982,08 | 1,01 | 988,08 | 1,07 | 1052,77 |

Як видно з наведених даних, ступінь коригування ваг домогосподарств по району, що розглядається, суттєво відрізняється за етапами. Найбільше ваги коригуються на етапі калібрації, що пов'язано з відносно малою кількістю домогосподарств, які обстежуються в сільській місцевості, і незначним рівнем відмов від обстеження.

3. ПОРЯДОК ВИЗНАЧЕННЯ ЯКОСТІ СИСТЕМИ ВАГ

Якість системи ваг ОУЖД визначається після кожного етапу їх коригування, а саме після урахування фактичної ймовірності відбору домогосподарств, відмов домогосподарств від обстеження, калібрації системи ваг (див. п. 2.2 – 2.4) окремо за результатами основного та всіх квартальних інтерв'ю. Розрахунки здійснюються на

національному рівні, тобто по всьому агрегованому масиву даних обстеження. За необхідності може бути здійснена перевірка якості системи ваг в межах окремих регіонів за типами місцевості.

При контролі якості системи ваг рекомендується розрахувати наступні характеристики.

1) Середня вага по всіх домогосподарствах як окремих множник – коефіцієнт інфляції $F_{(j)}$.

Величина $F_{(j)}$ для j -го етапу розрахунку ваг визначається за формулою:

$$F_{(j)} = \frac{\sum_{i=1}^n w_{(j)i}}{n} \quad (3.1)$$

Коефіцієнт інфляції є важливим параметром системи ваг, який показує скільки в середньому домогосподарств репрезентує одне обстежене домогосподарство на певному етапі ОУЖД.

2) Максимальне значення $w_{(j)\max}$ та мінімальне значення $w_{(j)\min}$ ваги домогосподарств. Визначається по всіх обстежених на певному етапі реалізації ОУЖД домогосподарствах.

При аналізі отриманих результатів необхідно звернути особливу увагу на те, що $w_{\min}^{(j)} \in 0$ є неприпустимим. За таких умов, як правило, необхідно перевірити процедури розрахунку системи ваг i , у разі необхідності, забезпечити додатні значення величин ваг із застосуванням спеціальних процедур [6].

Розраховуються також відношення $R_{(j)1}$, $R_{(j)2}$ та $R_{(j)3}$ за формулами:

$$R_{(j)1} = \frac{F_{(j)}}{w_{(j)\min}}; \quad R_{(j)2} = \frac{w_{(j)\max}}{F_{(j)}}; \quad R_{(j)3} = \frac{w_{(j)\max}}{w_{(j)\min}} \quad (3.2)$$

Ці відношення характеризують ступінь відхилення крайніх значень ваг від середнього значення та між собою

у відносних величинах. Значення $R_{(j)1}$ та $R_{(j)2}$ характеризують ступінь несиметричності розподілу ваг.

Ефективним способом зменшення величин $R_{(j)1}$, $R_{(j)2}$ та $R_{(j)3}$ є процедура зрізання ваг [10]. За цією процедурою ваги, що перевищують встановлені на основі спеціального статистичного аналізу максимальні

значення або є нижчими за встановлені мінімальні значення, замінюються на граничні значення: $w_i = w_{\min}$,

якщо $w_i < w_{\min}$, $w_i = w_{\max}$, якщо $w_i > w_{\max}$. Зазначений спеціальний статистичний аналіз є аналогічним аналізу первинних даних обстеження при виявленні викидних значень – „аутлайерів“.

3) Розмах варіації ваг за формулою:

$$\Delta R_{(j)} = w_{(j)\max} - w_{(j)\min} \quad (3.3)$$

Розмах варіації є характеристикою варіативності ваг.

4) Відносне зростання дисперсії вибірових оцінок, джерелом яких є варіативність ваг по одиницях вибірки, k_L . Величина k_L визначається за формулою [10]:

$$k_L = \frac{n \sum_{i=1}^n (w_{(j)i})^2}{(\sum_{i=1}^n w_{(j)i})^2} \quad (3.4)$$

Величина k_L характеризує ступінь зростання величини дисперсії вибірових оцінок показників ОУЖД за рахунок власної варіативності ваг. Це зростання еквівалентне зменшенню обсягу вибірки у k_L разів.

Ефективним способом зменшення величин k_L є процедура зрізання ваг.

5) Коефіцієнт варіації ваг $CV(w_{(j)})$ за формулою:

$$CV(w_{(j)}) = \sqrt{\frac{1}{n \cdot F_{(j)}^2} \cdot \sum_{i=1}^n (w_{(j)i} - F_{(j)})^2} \quad (3.5)$$

де $w_{\bar{i}}$ – вага \bar{i} -го домогосподарства (особи), $\bar{i} = 1, 2, \dots, n$.

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

$$R_{F(J)} = \frac{F_{(J)}}{F_{(J-1)}} \quad (3.6)$$

Порівняння коефіцієнтів інфляції для суміжних етапів розрахунку системи ваг дозволяє виявити етапи коригування ваг, які в середньому мають найбільший вплив на ваги домогосподарств (осіб).

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

$$R_{(J)A} = \frac{CV(w_{(J)})}{CV(w_{(J-1)})} \quad (3.7)$$

Порівняння коефіцієнтів варіації для суміжних етапів коригування ваг дає можливість виявити етапи, які в середньому мають найбільший вплив на дисперсію ваг домогосподарств (осіб).

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

$$r_{(J, J-1)} = \frac{\sum_{i=1}^n (w_{(J-1)i} - \bar{w}_{(J-1)}) \cdot (w_{(J)i} - \bar{w}_{(J)})}{\sqrt{\sum_{i=1}^n (w_{(J-1)i} - \bar{w}_{(J-1)})^2 \cdot \sum_{i=1}^n (w_{(J)i} - \bar{w}_{(J)})^2}} \quad (3.8)$$

Коефіцієнт кореляції (3.8) характеризує щільність лінійного статистичного зв'язку систем 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.

Прийнятність розрахованої системи ваг визначається на основі отриманих значень характеристик якості 1) – 8) шляхом їх порівняння з встановленими граничними значеннями, характеристиками якості за попередні періоди ОУЖД та на основі консультацій з експертами. Необхідно зазначити, що прийняті граничні значення характеристик якості системи ваг залежать від результатів обстежень і тому для кожного року проведення ОУЖД визначаються окремо з урахуванням особливостей поточної ситуації.

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

$$N_{(J)} = \sum_{i=1}^n w_{(J)i} \quad (3.9)$$

Оцінка загальної кількості домогосподарств $N_{(J)}$ порівнюється з оцінкою загальної кількості домогосподарств з зовнішніх джерел НЗІ (попередньою оцінкою загальної кількості домогосподарств, що використовувалась при побудові вибірки, або з іншою достовірною інформацією щодо загальної кількості неінституційних домогосподарств в Україні). Розраховується абсолютне відхилення за формулою:

$$D_{(J)} = N_{\text{НЗІ}} - N_{(J)} \quad (3.10)$$

Отримані результати аналізуються на підставі існуючих даних щодо рівня надійності вибірки, зокрема перевіряється чи є відхилення меншим, ніж довірчий інтервал для точкової оцінки $N_{(J)}$.

10) По масиву первинних даних для осіб, що проживають в домогосподарствах, по яких проведено інтерв'ю,

з використанням ваг $w_{(J)i}$, розраховуються оцінки загальної чисельності населення України та чисельності населення за розрізами регіон – тип місцевості. Результати порівнюються з даними демографічної статистики.

Приклад результатів оцінювання якості кінцевої системи ваг наведено у табл. 3.1. Наведені дані ілюструють зміну окремих характеристик якості системи ваг ОУЖД протягом 1999 – 2004 років.

Як видно з наведених даних, протягом 1999 – 2004 років спостерігається тенденція до зменшення величини середньої ваги. Враховуючи, що обсяг щорічної вибірки ОУЖД протягом 1999 – 2004 років практично не змінювався, це можна пояснити, головним чином, зменшенням кількості домогосподарств у генеральній сукупності. В обстеженні також збільшується коефіцієнт варіації ваг, але несуттєво. Величини інших характеристик в обстеженні умов життя домогосподарств свідчать про погіршення якості системи ваг. Це пояснюється поетапним впровадженням у 2001 і 2002 роках процедури калібрації ваг, що призначена для коригування системи ваг у відповідності з наявною зовнішньою інформацією і, як правило, призводить до більш суттєвих змін величини ваг, що підтверджується також і міжнародним досвідом застосування процедур калібрації.

Таблиця 3.1.

Зміна окремих характеристик якості кінцевої системи ваг в ОУЖД за 1999-2004 роки

| Характеристики | Рік | | | | | |
|--------------------------------------|-------|-------|-------|--------|--------|--------|
| | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 |
| Середня вага, \bar{F} | 1921 | 1897 | 1889 | 1868 | 1812 | 1753 |
| Мінімальна вага, w_{\min} | 591 | 384 | 357 | 46 | 45 | 56 |
| Максимальна вага, w_{\max} | 5585 | 10205 | 9456 | 7195 | 13817 | 7333 |
| Коефіцієнт варіації ваг, $CV(w)$, % | 31,40 | 34,96 | 37,16 | 38,50 | 46,08 | 44,97 |
| = | 3,25 | 4,94 | 5,29 | 40,63 | 40,28 | 30,30 |
| = | 2,91 | 5,38 | 5,00 | 3,85 | 7,62 | 4,18 |
| = | 9,45 | 26,58 | 26,49 | 156,41 | 307,04 | 130,95 |
| Розмах варіації ваг, | 4994 | 9821 | 9099 | 7149 | 13772 | 7277 |

Слід зазначити також, що після впровадження у 2004 році нової системи організації ОУЖД якість статистичних ваг покращилась у порівнянні з результатами 2003 року, про що свідчать значення всіх наведених характеристик якості.

4. ІНФОРМАЦІЙНА БАЗА РОЗРАХУНКУ СИСТЕМИ СТАТИСТИЧНИХ ВАГ ОУЖД

Інформаційна база для реалізації процедур розрахунку системи ваг за результатами основного та чотирьох кварталних інтерв'ю складається з трьох основних частин:

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

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 ступеню (міських поселень та сільських рад), ;
- імовірність відбору одиниць другого ступеня (інструкторські дільниці – по міських поселеннях,

домогосподарства - по сільській місцевості), ;

- імовірність відбору домогосподарств по міських поселеннях, ;

- загальна кількість домогосподарств у ВТОВ по міських поселеннях, ;

- загальна кількість домогосподарств у ПТОВ по сільській місцевості, ;

- фактична кількість відібраних для обстеження домогосподарств у територіальній одиниці вибірки (у ВТОВ

- по міських поселеннях, у ПТОВ – по сільській місцевості), ;

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

Результати ОУЖД включають наступні дані:

- масив даних по обстежених домогосподарствах за результатами основного та кожного кварталного інтерв'ю з наступними ознаками:

- код домогосподарства;
- загальнодержавний номер територіальної одиниці (ВТОВ по міських поселеннях, ПТОВ по сільській

місцевості);

- код регіону;

- код економічного району;
- тип місцевості;
- тип домогосподарства за розміром;
- кількість осіб в домогосподарстві;
- ознака наявності в домогосподарстві дітей;
- кількість жінок віком 0-17 років в домогосподарстві;
- кількість жінок віком 18-35 років в домогосподарстві;
- кількість жінок віком 36-54 років в домогосподарстві;
- кількість жінок віком старше 54 років в домогосподарстві;
- кількість чоловіків віком 0-17 років в домогосподарстві;
- кількість чоловіків віком 18-35 років в домогосподарстві;
- кількість чоловіків віком 36-59 років в домогосподарстві;
- кількість чоловіків віком старше 59 років в домогосподарстві;
- додаткові ознаки:
- номер комірки для врахування відмов домогосподарств від обстеження, I;
- код комірки, ;

- загальна початкова кількість домогосподарств у комірці, .
 Структура інформації, яка формується за результатами ОУЖД, залежить від особливостей реалізації процедури врахування відмов домогосподарств від обстеження та процедури калібрації системи статистичних ваг. Інформація з зовнішніх джерел містить наступні дані: за регіонами

- загальна кількість неінституційних домогосподарств великих міст;
- загальна кількість неінституційних домогосподарств малих міст;
- загальна кількість неінституційних домогосподарств сільської місцевості;
- чисельність неінституційного населення великих міст;
- чисельність неінституційного населення малих міст;
- чисельність неінституційного населення сільської місцевості;
- загальна кількість неінституційних домогосподарств з дітьми;
- чисельність жінок віком до 18 років;
- чисельність жінок віком 18-35 років;
- чисельність жінок віком 36-54 роки;
- чисельність жінок віком старше 54 років;
- чисельність чоловіків віком до 18 років;
- чисельність чоловіків віком 18-35 років;
- чисельність чоловіків віком 36-59 років;
- чисельність чоловіків віком старше 59 років.

Масив зовнішньої інформації створюються один раз на рік перед розрахунком системи статистичних ваг домогосподарств за результатами основного інтерв'ю ОУЖД.

Основними джерелами зовнішньої інформації є:

- 1) дані демографічної статистики ("Чисельність наявного населення України", "Склад населення України за статтю та віком", станом на 1 січня кожного року);
- 2) дані соціальної статистики;
- 3) експертні оцінки щодо середнього розміру домогосподарств за типами місцевості та регіонами та кількості домогосподарств з дітьми за регіонами. Для побудови експертних оцінок використовуються дані інших вибірових обстежень населення, дані ОУЖД попередніх років, інша наявна інформація.

ДЖЕРЕЛА

1. Вибіркове спостереження: Термінологічний словник. – К.: НТК статистичних досліджень, 2004.
2. Kish L. Survey sampling. – New York: Wiley, 1995.
3. Kalton G. Introduction to Survey Sampling. – Beverly Hills: Sage, 1983.
4. Доходи і витрати домогосподарств України у 2004 році: Стат. збірник. – К.: Державний комітет статистики України, 2005.
5. 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.
6. Саріогло В.Г. Проблеми статистичного зважування вибірових даних. – К.: ІВЦ Держкомстату України, 2005.
7. Deville J.-C., Sarndal C.-E. Calibration Estimators in Survey Sampling // Journal of the American Statistical Association. – 1992. – Vol. 87, № 418.
8. Deville J.-C., Sarndal C.-E., Sautory O. Generalized Raking Procedures in Survey Sampling // Journal of the American Statistical Association. – 1993. – Vol. 88, № 423.
9. Україна: прогноз розвитку продуктивних сил. – К.: РВПС України НАН України, 1998. – Т. 2. Додатки.
10. Kish L.: Selected papers. – New York: Wiley, 2003.
11. Lundstrom S., Sarndal C.-E. Estimation in the Presence of Nonresponse. – Stockholm: Statistics Sweden, 2001.

ДОДАТОК А

Зовнішня інформація для калібрації системи статистичних ваг у 2005 році

Таблиця А.1

Зовнішня інформація щодо кількості домогосподарств за типами місцевості станом на 1.01.2005

| | Всього | в тому числі: | | |
|-------------------|-----------------|------------------|----------------|------------------------|
| | | у великих містах | у малих містах | у сільській місцевості |
| Україна | 17539366 | 7543997 | 4465540 | 5529829 |
| АР Крим | 748967 | 360932 | 131715 | 256320 |
| Вінницька | 677950 | 151466 | 153492 | 372992 |
| Волинська | 343650 | 73885 | 103361 | 166404 |
| Дніпропетровська | 1355104 | 904651 | 229642 | 220811 |
| Донецька | 1751670 | 1046075 | 538793 | 166802 |
| Житомирська | 512740 | 107320 | 172971 | 232449 |
| Закарпатська | 363542 | 38708 | 108996 | 215838 |
| Запорізька | 712465 | 398566 | 143317 | 170582 |
| Івано-Франківська | 462012 | 82262 | 124126 | 255624 |
| Київська | 666114 | 87767 | 283534 | 294813 |
| Кіровоградська | 447392 | 142371 | 127178 | 177843 |
| Луганська | 966099 | 555121 | 282967 | 128011 |
| Львівська | 835701 | 278253 | 246323 | 311125 |
| Миколаївська | 465991 | 195662 | 115861 | 154468 |
| Одеська | 854301 | 368313 | 206095 | 279893 |
| Полтавська | 634484 | 214013 | 158084 | 262387 |
| Рівненська | 386320 | 80365 | 109811 | 196144 |
| Сумська | 474745 | 106525 | 198931 | 169289 |
| Тернопільська | 372793 | 75055 | 91747 | 205991 |
| Харківська | 1058799 | 541569 | 297314 | 219916 |
| Херсонська | 421828 | 145089 | 115408 | 161331 |
| Хмельницька | 513734 | 100279 | 157632 | 255823 |
| Черкаська | 548646 | 125530 | 164906 | 258210 |
| Чернівецька | 311217 | 94522 | 40147 | 176548 |
| Чернігівська | 486581 | 110040 | 163189 | 213352 |
| м. Київ | 1032442 | 1032442 | 0 | 0 |
| м. Севастополь | 134079 | 127216 | 0 | 6863 |

Зовнішня інформація щодо чисельності неінституціонального населення за типами місцевості станом на 1.01.2005

| | Всього | в тому числі: | | |
|-------------------|-----------------|------------------|-----------------|------------------------|
| | | у великих містах | у малих містах | у сільській місцевості |
| Україна | 45830345 | 19127181 | 11516660 | 15186504 |
| АР Крим | 1968115 | 863415 | 360927 | 743773 |
| Вінницька | 1704963 | 425162 | 365623 | 914178 |
| Волинська | 1033672 | 182722 | 329026 | 521924 |
| Дніпропетровська | 3322616 | 2172438 | 586485 | 563693 |
| Донецька | 4346977 | 2592437 | 1324744 | 429796 |
| Житомирська | 1325176 | 303477 | 434599 | 587100 |
| Закарпатська | 1239840 | 107036 | 345181 | 787623 |
| Запорізька | 1798975 | 993088 | 366377 | 439510 |
| Івано-Франківська | 1383812 | 213969 | 363120 | 806723 |
| Київська | 1757017 | 217595 | 800875 | 738547 |
| Кіровоградська | 1068386 | 367917 | 276378 | 424091 |
| Луганська | 2346811 | 1344338 | 675951 | 326522 |
| Львівська | 2546510 | 816955 | 690452 | 1039103 |
| Миколаївська | 1214745 | 503793 | 302192 | 408760 |
| Одеська | 2278419 | 984857 | 502432 | 791130 |
| Полтавська | 1551756 | 525869 | 383306 | 642581 |
| Рівненська | 1151143 | 227877 | 304958 | 618308 |
| Сумська | 1198664 | 299355 | 480411 | 418898 |
| Тернопільська | 1109835 | 203509 | 263696 | 642630 |
| Харківська | 2636147 | 1318153 | 756050 | 561944 |
| Херсонська | 1123727 | 383736 | 287297 | 452694 |
| Хмельницька | 1373161 | 302456 | 402927 | 667778 |
| Черкаська | 1342597 | 331865 | 392706 | 618026 |
| Чернівецька | 901194 | 242296 | 118485 | 540413 |
| Чернігівська | 1167843 | 284287 | 402462 | 481094 |
| м. Київ | 2609301 | 2609301 | 0 | 0 |
| м. Севастополь | 328943 | 309278 | 0 | 19665 |

Зовнішня інформація щодо кількості домогосподарств з дітьми за типами місцевості станом на 1.01.2005

| | Всього |
|-------------------|----------------|
| Україна | 6577249 |
| АР Крим | 274782 |
| Вінницька | 232181 |
| Волинська | 167302 |
| Дніпропетровська | 466900 |
| Донецька | 611434 |
| Житомирська | 189335 |
| Закарпатська | 190072 |
| Запорізька | 253179 |
| Івано-Франківська | 215735 |
| Київська | 244172 |
| Кіровоградська | 145712 |
| Луганська | 336246 |
| Львівська | 374201 |
| Миколаївська | 180247 |
| Одеська | 338809 |
| Полтавська | 209061 |
| Рівненська | 181623 |
| Сумська | 158846 |
| Тернопільська | 170987 |
| Харківська | 356401 |
| Херсонська | 156403 |
| Хмельницька | 199295 |
| Черкаська | 192304 |
| Чернівецька | 159546 |
| Чернігівська | 165023 |
| м. Київ | 357633 |
| м. Севастополь | 49820 |

Таблиця А.4

Зовнішня інформація щодо розподілу населення за статеві-віковими групами станом на 1.01.2005

| | Чоловіки у віці (років) | | | | Жінки у віці (років) | | | |
|-------------------|-------------------------|----------------|----------------|----------------|----------------------|----------------|----------------|----------------|
| | до 18 | 19-35 | 36-59 | 60 і старші | до 18 | 19-35 | 36-54 | 55 і старші |
| Україна | 4598579 | 6169727 | 7009751 | 3380872 | 4339470 | 6141686 | 6687341 | 7502919 |
| АР Крим | 193583 | 263288 | 310261 | 136337 | 181995 | 262796 | 309098 | 310757 |
| Вінницька | 179729 | 210976 | 249929 | 138639 | 167751 | 215862 | 230274 | 311803 |
| Волинська | 126816 | 141174 | 150215 | 67726 | 118866 | 141615 | 136549 | 150711 |
| Дніпропетровська | 316357 | 448114 | 503837 | 253492 | 294568 | 446268 | 501922 | 558058 |
| Донецька | 379095 | 573131 | 686121 | 343948 | 359758 | 574838 | 669538 | 760548 |
| Житомирська | 143827 | 169184 | 196750 | 103901 | 135452 | 168009 | 179003 | 229050 |
| Закарпатська | 158168 | 179337 | 186515 | 72449 | 148980 | 173375 | 168722 | 152294 |
| Запорізька | 169812 | 237025 | 281738 | 135461 | 157269 | 238502 | 277146 | 302022 |
| Івано-Франківська | 163998 | 189142 | 205486 | 94637 | 153947 | 186468 | 186919 | 203215 |
| Київська | 174653 | 233243 | 269879 | 132315 | 164119 | 229449 | 254871 | 298488 |
| Кіровоградська | 108615 | 130972 | 161693 | 88146 | 101524 | 134585 | 152883 | 189968 |
| Луганська | 207227 | 311687 | 374788 | 181242 | 193108 | 309106 | 361514 | 408139 |
| Львівська | 280824 | 358619 | 388914 | 178794 | 268889 | 339723 | 351522 | 379225 |
| Миколаївська | 124612 | 163552 | 188798 | 85569 | 116790 | 162598 | 182316 | 190510 |
| Одеська | 230828 | 323097 | 345618 | 164583 | 219966 | 310902 | 335032 | 348393 |
| Полтавська | 148124 | 198187 | 235166 | 127780 | 138617 | 198110 | 222005 | 283767 |
| Рівненська | 146903 | 159570 | 166215 | 71908 | 138245 | 158744 | 150903 | 158653 |
| Сумська | 112183 | 150593 | 189887 | 95231 | 104441 | 153268 | 177506 | 215555 |
| Тернопільська | 122789 | 144949 | 163555 | 82666 | 116601 | 148646 | 147711 | 182918 |
| Харківська | 230361 | 378858 | 410137 | 190542 | 217759 | 365025 | 400797 | 442668 |
| Херсонська | 119823 | 150140 | 174600 | 78074 | 111332 | 149370 | 166415 | 173973 |
| Хмельницька | 144890 | 172579 | 208019 | 107644 | 136275 | 169937 | 190973 | 242844 |
| Черкаська | 132283 | 164700 | 202030 | 111973 | 123820 | 168863 | 190752 | 248176 |
| Чернівецька | 119891 | 107460 | 127680 | 53046 | 130409 | 125822 | 116973 | 119913 |
| Чернігівська | 106982 | 142038 | 177979 | 102139 | 99394 | 138844 | 163601 | 236867 |
| м. Київ | 227717 | 419658 | 403741 | 159704 | 212922 | 426353 | 409765 | 349441 |
| м. Севастополь | 28489 | 48454 | 50200 | 22926 | 26673 | 44608 | 52631 | 54963 |