

Методика розрахунку системи статистичних ваг для оцінювання показників вибіркового обстеження сільськогосподарської діяльності населення в сільській місцевості

Методика розрахунку системи статистичних ваг для оцінювання показників вибіркового обстеження сільськогосподарської діяльності населення у сільській місцевості (далі – Методика) розроблена відповідно до Плану заходів щодо реалізації у 2010 році Стратегії розвитку державної статистики на період до 2012 року і визначає методи оцінювання показників вибіркового обстеження сільськогосподарської діяльності населення в сільській місцевості, що проводяться органами державної статистики, з урахуванням особливостей дизайну та проведення цього обстеження. Вона містить опис як теоретичних основ, так практичних кроків застосування обраних методів отримання оцінок показників згаданого обстеження, а також показників щодо контролю якості статистичних ваг.

Методика призначена для використання фахівцями органів державної статистики на державному та регіональному рівні, науково-дослідних установ, інших зацікавлених користувачів статистичної інформації.

Передмова

На виконання Указу Президента України від 3 грудня 1999 року № 1529 "Про невідкладні заходи щодо прискорення реформування аграрного сектора економіки", з метою вдосконалення статистичного обліку виробництва сільськогосподарської продукції господарствами населення у вересні 2000 року було запроваджене вибіркове обстеження сільськогосподарської діяльності домогосподарств у сільській місцевості (далі – обстеження СГД). Програма згаданого статистичного спостереження є комплексною. Дані, що отримуються завдяки цьому обстеженню, є інформаційною базою для формування основних показників сільськогосподарського виробництва за такою специфічною категорією товаровиробників, як господарства населення. Вона охоплює різнопланові показники, які стосуються сільськогосподарського виробництва та об'єднані у два запитальники: базового і щомісячного інтерв'ю. Це такі показники, як наявність і використання земельних ділянок у домогосподарствах, структура посіяних площ, чисельність та видовий склад сільськогосподарських тварин, наявність техніки та господарських споруд, залучення найманої праці, кількість отриманої в домогосподарстві продукції рослинництва і тваринництва, кількість продукції власного виробництва, що продана домогосподарством, площі посівів озимих культур та площі загиблих посівів, наявність залишків основних продуктів у домогосподарстві.

Такий широкий діапазон показників, а також неоднорідність сукупності домогосподарств, що вивчається, необхідність отримання якісних результатів обстеження на рівні регіонів визначили складний дизайн обстеження сільськогосподарської діяльності населення і зумовили потребу в застосуванні більш складних методів екстраполяції даних обстеження на генеральну сукупність, ніж, наприклад, метод середніх величин.

Вибір методів поширення результатів вибіркового обстеження сільськогосподарської діяльності населення, які подані в цій Методіці, був зумовлений також і тим, що при побудові вибірки в якості характеристики одиниці відбору використовуються показники, що є основоположними у здійсненні сільськогосподарської діяльності, а саме розмір площ земельних ділянок, якими користуються домогосподарства. Також по відібраних територіальних одиницях здійснюється розшарування сукупності домогосподарств за розміром землі з відповідним обсягом вибірки в кожній страті.

Ця Методика є однією з основних складових методичного забезпечення проведення вибіркового обстеження сільськогосподарської діяльності населення. Процедури і алгоритми, передбачені нею, орієнтовані на використання сучасних технологій обробки статистичної інформації.

1. Визначення термінів

У цій Методіці терміни використані в такому значенні:

Відбір з імовірністю, 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 установленого 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 генератора 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 включеною до вибірки.

Відбір багатоступеневий – відбір, який відбувається у кілька послідовних ступенів (етапів): на першому ступені за основи вибірки відбираються певні відносно великі одиниці; на другому ступені всередині кожної відібраної 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 несудового спостереження, при якому обстежується не всі одиниці сукупності, а лише певним чином відібрана їх частина, а одиницю 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).

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 принципів її формування. Якістю плану вибірки значною мірою визначається репрезентативність вибіркової сукупності, отриманої за визначеними у плані правилами.

Імовірність – 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 взяли 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 вибіркової сукупності. Необхідний для забезпечення репрезентативності та належної якості результатів вибіркового обстеження обсяг вибірки розраховують залежно від способу відбору та типу вибірки. Величина обсягу вибірки залежить від варіації досліджуваних ознак і припустимої похибки вибірки.

Одиниця відбору – одиниця генеральної сукупності чи її 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 – первинний елемент об'єкта статистичного спостереження, який є носієм ознак, що підлягають реєстрації у ході статистичного спостереження. Вибір одиниці спостереження залежить від мети та конкретних умов проведення спостереження.

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

Репрезентативність – властивість вибіркової сукупності повно й адекватно відобразити основні характеристики генеральної сукупності.

Сільські домогосподарства – домогосподарства, які мають реєстрацію місця проживання на території сільського населеного пункту.

Статистична вага – 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, який підпорядковані сільські населені пункти.

2. Принципи оцінювання показників вибіркового обстеження сільськогосподарської діяльності населення в сільській місцевості

Обстеження СГД є вибіркового обстеження складного дизайну і тому для адекватного і повного відображення явища, що вивчаються в генеральній сукупності за його допомогою, будується система статистичних ваг. Формування системи статистичних ваг є одним із основних етапів обробки й поширення на генеральну сукупність результатів будь-якого вибіркового обстеження, в тому числі й обстеження сільськогосподарської діяльності населення.

Важливість цього етапу зумовлена тим, що некоректна система ваг знижує якість оцінок показників для генеральної сукупності, розрахованих за даними вибіркового обстеження. Статистичне зважування підвищує рівень репрезентативності вибірки в плані відображення нею кількісних і структурних характеристик генеральної сукупності та забезпечує можливість вирішення таких проблем, як урахування нерівних ймовірностей відбору одиниць вибірки, фактичних рівнів відомо домогосподарств від участі в обстеженні, узгодження результатів обстеження з достовірною зовнішньою інформацією. Таким чином, система статистичних ваг відображає дизайн вибірки, а також урахує основні етапи реалізації обстеження.

Генеральну сукупність обстеження СГД складають домогосподарства сільських населених пунктів, підпорядкованих сільським, селищним, міським радам (далі – сільради), місце проживання яких зареєстроване на їхній території ("постійно проживаючі") й у володінні та користуванні яких є земельні ділянки, отримані ними відповідно до норм чинного законодавства. Інформаційною базою генеральної сукупності цього обстеження є дані погосподарського обліку.

Дизайн вибірки для обстеження СГД розроблено у відповідності з сучасною міжнародною практикою проведення вибіркового обстежень [3, 5]. Формування вибірки здійснюється на основі процедури стратифікованого багатоступеневого ймовірнісного (випадкового) відбору у чотири основні етапи [7]:

- 1) виключення територій, що не можуть бути обстежені;
- 2) стратифікація генеральної сукупності з урахуванням адміністративно-територіального устрою України;
- 3) відбір територіальних одиниць з урахуванням природно-сільськогосподарського районування;
- 4) відбір домогосподарств.

Основною умовою формування вибірки є обов'язкове представлення у ній Автономної Республіки Крим, усіх областей і адміністративних районів сільської місцевості, за виключенням територій, які належать до першої або другої зони радіоактивного забруднення внаслідок аварії на Чорнобильській АЕС (зона відчуження та зона безумовного (обов'язкового) відселення).

З метою адекватного відображення у вибірці основних особливостей адміністративно-територіального поділу України, генеральна сукупність домогосподарств стратифікується за регіонами, а в межах регіонів – за адміністративними районами. В межах кожної страти проводиться двоступенева процедура відбору домогосподарств, які розбиті на чотири однорідні за основними характеристиками групи: малі домогосподарства, середні домогосподарства, великі домогосподарства і дуже великі домогосподарства (у межах територіальної одиниці вибірки).

Відбір як територіальних одиниць, так і домогосподарств здійснюється на основі механізму систематичного відбору одиниць з імовірністю, пропорційною площі землі, яка знаходиться у володінні та користуванні домогосподарств (далі – площа землі).

На першому етапі для проведення обстеження СГД у кожному адміністративному районі сільської місцевості відбирається не менше ніж три первинні територіальні одиниці вибірки (ПТОВ) – сільради або їхні групи. При цьому враховується поділ країни на природно-сільськогосподарській території (далі – ПСТ). У районі, віднесеному до більше, ніж однієї ПСТ, кожна ПСТ представляється у вибірці хоча б однією територіальною одиницею. У кожному регіоні відбирається щонайменше одна територіальна одиниця вибірки, сформована з сільських населених пунктів, підпорядкованих міським або селищним радам районного значення.

На другому етапі в межах ПТОВ відбираються домогосподарства.

Це є принципова схема побудови вибірки для обстеження СГД.

У ймовірній вибірці обстеження СГД кожне i -те домогосподарство генеральної сукупності обсягу N , $i = 1, 2, \dots, N$ повинне мати певну позитивну ймовірність P_i потрапити до вибірки. Тобто, за результатами відбору це домогосподарство своєю площею землі буде репрезентувати певну площу землі одиниць генеральної сукупності i , відповідно, при оцінюванні

показників у генеральній сукупності його потрібно врахувати $1/p_i$ разів. Таким чином, статистична вага одиниці вибірки при поширенні результатів обстеження на генеральну сукупність розраховується за такою формулою:

$$w_{li} = \frac{1}{p_i} \quad (2.1)$$

Імовірність P_i є загальною ймовірністю включення домогосподарства до вибірки, а вага w_{li} , розрахована на її основі, є базовою вагою. Базова вага визначається дизайном вибірки [3, 5, 9].

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:

$$w_i = w_{li} \cdot k_{1i} \cdot k_{2i} \cdot \dots \cdot k_{Ei}, \quad (2.2)$$

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

Для обстеження СГД розрахунок системи статистичних ваг складається з таких основних етапів:

- розрахунок базових ваг домогосподарств;
- коригування ваг для врахування руху площі землі домогосподарств у ПТОВ;
- коригування ваг для врахування ступеня участі відібраних одиниць в обстеженні;
- коригування ваг для узгодження результатів обстеження з наявною зовнішньою інформацією;
- контроль якості системи ваг.

Оскільки базові ваги враховують реальний дизайн вибірки (принципи побудови, структуру вибірки, механізми відбору тощо), то етап розрахунку системи базових ваг домогосподарств є визначальним.

Базова вага, що визначається за формулою (2.1), – це результат певної послідовності розрахунків. Існуючий дизайн вибірки обстеження СГД зумовлює використання для побудови базових ваг таких складових:

– імовірність відбору ПТОВ P_{li} ;

– імовірність відбору домогосподарств у межах ПТОВ P_{2i} .

Базовою вагою i -го домогосподарства є обернена величина загальної імовірності відбору домогосподарства:

$$w_{li} = \frac{1}{p_i} = \frac{1}{P_{li}} \cdot \frac{1}{P_{2i}} \quad (2.3)$$

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 домогосподарств, домогосподарств, що тимчасово вибувають з обстеження, або з якими неможливо встановити контакт, та тих домогосподарств, які відмовились від участі в обстеженні, а також зміни в площах землі домогосподарств.

Одним з найбільш розповсюджених і теоретично обґрунтованих підходів до 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.

Для врахування змін у площах землі домогосподарств проводиться актуалізація основних характеристик генеральної та вибіркової сукупностей. Базові ваги домогосподарств коригуються на коефіцієнт, який відображає відношення актуалізованих значень площі землі домогосподарств до тих, які використовувались при процедурі відбору.

Для узгодження результатів вибіркового обстеження з наявною зовнішньою інформацією щодо певних характеристик генеральної сукупності здійснюється спеціальне коригування системи статистичних ваг. Доцільність цього зумовлена такими чинниками:

1) оцінки показників за результатами вибіркового обстеження завжди характеризуються певною похибкою, джерелами якої є несучільність обстеження (похибка вибірки);

2) вибіркоче обстеження не може забезпечити оцінок певних характеристик генеральної сукупності, наприклад оцінок загальної кількості її одиниць, загальної площі землі. Хоча ці оцінки й отримуються за результатами обстеження, вони відображають головним чином параметри, що були закладені при формуванні вибірки. Коригування статистичних ваг здійснюється 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.

З відомих на сьогодні узагальнених методів коригування системи статистичних ваг найбільш теоретично розроблено й ефективно є так звана процедура калібрації статистичних ваг [1, 2, 9].

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.

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

– може погіршуватись якість системи ваг, зокрема, для окремих домогосподарств із 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.

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

Таким чином, з урахуванням наведеного вище загальна формула, за якою розраховується система статистичних ваг домогосподарств обстеження СГД, має такий вигляд:

$$w_i = \frac{1}{p_i} = \frac{1}{P_{li}} \cdot \frac{1}{P_{2i}} \cdot k_i^{(зем)} \cdot k_i^{(ч)} \cdot k_i^{(калібр)} \quad (2.4)$$

3. Порядок розрахунку складових системи ваг

3.1. Розрахунок базової ваги домогосподарств

3.1.1. Визначення ймовірності відбору ПТОВ P_{li}

Інформаційною базою для визначення ймовірності відбору ПТОВ є результати формування вибірки, які включають основу вибірки, кількість відібраних ПТОВ в g -ому районі, ПСТ

(d_r) та їх перелік, інформацію щодо розміру площі землі домогосподарств району, ПСТ (S_r) та інформацію щодо розміру площі землі по кожній відібраній ПТОВ (S_{rj}) . Імовірність відбору та базові ваги визначаються при формуванні вибірки один раз на п'ять років.

Відбір саморепрезентативних ПТОВ

На першому етапі формування вибірки ПТОВ визначаються саморепрезентативні ПТОВ [7]. Для цього у межах кожного району, ПСТ визначають порогове значення площі землі для

відібраних ПТОВ $(S_{пор,r}^{(I)})$:

$$S_{пор,r}^{(I)} = \frac{S_r}{d_r}, \quad (3.1)$$

де $r = 1, 2, \dots, R$, де R – загальна кількість районів, ПСТ, по яких відбираються ПТОВ.

Якщо площа землі ПТОВ більше або дорівнює пороговому значенню $S_{пор,r}^{*(I)}$, така ПТОВ вважається саморепрезентативною і для всіх домогосподарств цієї ПТОВ імовірність відбору

на першому етапі дорівнює одиниці, $P_{li} = 1$.

Таким чином, саморепрезентативні ПТОВ уключаються до вибірки без процедури відбору.

Далі саморепрезентативні ПТОВ виключаються з переліку ПТОВ району, ПСТ, а їхня площа – з площі землі району, ПСТ. Таким чином, кількість ПТОВ, яку необхідно відібрати,

дорівнюватиме $d_r - d_r^{*(I)}$, де $d_r^{*(I)}$ – кількість саморепрезентативних ПТОВ r -ого району, ПСТ, а площа землі району, ПСТ без площі саморепрезентативних ПТОВ складає $S_r - S_r^{*(I)}$, де $S_r^{*(I)}$ – їх загальна площа землі.

Після цього здійснюється визначення нового порогового значення саморепрезентативності ПТОВ $S_{пор,r}^{*(I)}$ за формулою:

$$S_{пор,r}^{*(I)} = \frac{S_r - S_r^{*(I)}}{d_r - d_r^{*(I)}}, \quad (3.2)$$

ПТОВ, для яких площа землі більше або дорівнює пороговому значенню $S_{пор,r}^{*(I)}$, також є саморепрезентативними.

Ця процедура може повторюватись необхідну кількість разів, доки не будуть виявлені всі саморепрезентативні ПТОВ у районі, ПСТ.

Імовірність відбору несаморепрезентативних ПТОВ визначається за формулами [7]:

якщо в районі, ПСТ немає саморепрезентативних ПТОВ:

$$p_{li} = d_r \cdot \frac{S_{r,l}}{S_r}, \quad (3.3)$$

де $S_{r,l}$ – площа землі l -ої ПТОВ r -ого району, ПСТ;

якщо в районі виявлено d_r^{**} саморепрезентативних ПТОВ:

$$p_{li} = (d_r - d_r^{**}) \cdot \frac{S_{r,l}}{S_r - S_r^{**}}, \quad (3.4)$$

де d_r^{**} – загальна кількість саморепрезентативних ПТОВ у r -ому районі, ПСТ, S_r^{**} – їх загальна площа землі.

3.1.2. Визначення ймовірності відбору домогосподарств у межах ПТОВ P_{Σ}

Стратифікація сукупності домогосподарств

У відповідності з методикою формування вибірки стратифікація домогосподарств у межах відібраних ПТОВ здійснюється з метою більш адекватного охоплення вибіркою домогосподарств із урахуванням двох чинників: розміру площі землі, якою користується домогосподарство, та обсягу сільськогосподарської продукції, виробленої у домогосподарстві [7].

Усі домогосподарства розподіляються за площею землі на 4 страти:

I страта – домогосподарства з площею землі до 0,5 га включно;

II страта – домогосподарства з площею землі від 0,51 га до 1,0 га включно;

III страта – домогосподарства з площею землі понад 1,01 га;

IV страта – домогосподарства, площа землі яких є більшою за поріг саморепрезентативності домогосподарств для певної ПТОВ. Такі домогосподарства є саморепрезентативними і обов'язково включаються до вибіркової сукупності домогосподарств ПТОВ.

Для кожної з відібраних ПТОВ розраховується поріг саморепрезентативності домогосподарств.

Поріг саморепрезентативності для домогосподарств l -ої ПТОВ r -го району визначається шляхом ділення загальної площі землі домогосподарств цієї ПТОВ $S_{r,l}$ на кількість домогосподарств, яку необхідно обстежити у цій ПТОВ, $n_{r,l}$:

$$S_{пор,r,l}^{(II)} = \frac{S_{r,l}}{n_{r,l}}. \quad (3.5)$$

Якщо площа землі домогосподарства більше або дорівнює пороговому значенню $S_{пор,r,l}^{(II)}$, таке домогосподарство вважається саморепрезентативним і для нього ймовірність

відбору дорівнює одиниці, $P_{\Sigma} = 1$.

Таким чином, саморепрезентативні домогосподарства включаються до вибірки без процедури відбору.

Далі саморепрезентативні домогосподарства виключаються з переліку домогосподарств ПТОВ, а їх площа – з площі землі ПТОВ. Таким чином, кількість домогосподарств, яку

необхідно відібрати, дорівнює $n_{r,l} - n_{r,l}^{*(II)}$, де $n_{r,l}^{*(II)}$ – кількість саморепрезентативних домогосподарств r -го району l -ої ПТОВ, а площа землі ПТОВ без площі саморепрезентативних

домогосподарств складає $S_{r,l} - S_{r,l}^{*(II)}$, де $S_{r,l}^{*(II)}$ – їх загальна площа землі.

$$S_{пор,r,l}^{*(II)} = \frac{S_{r,l} - S_{r,l}^{*(II)}}{n_{r,l} - n_{r,l}^{*(II)}}, \quad (3.6)$$

Домогосподарства, для яких площа землі більше або дорівнює пороговому значенню $S_{пор,r,l}^{*(II)}$, є саморепрезентативними.

Ця процедура може повторюватись необхідну кількість разів, доки не будуть виявлені всі саморепрезентативні домогосподарства у ПТОВ.

Визначення ймовірності відбору домогосподарств у межах ПТОВ P_{Σ}

Розрахунок імовірності відбору домогосподарства в межах ПТОВ здійснюється з урахуванням стратифікації та проводиться для домогосподарств у кожній страті.

При цьому враховується, що ПТОВ можуть бути як "повними" – усі страти ПТОВ представлені звичайними домогосподарствами, тобто такими, які погодилися брати участь у обстеженні, так і "неповними" – є страта, в якій немає звичайних домогосподарств, а присутні лише домогосподарства, які тимчасово вибувають з обстеження, неконтактні, зачинені або ті, які відмовились брати участь у обстеженні чи вибули в середині року через переїзд або смерть господаря.

Аналогічно, якщо в страті, окрім звичайних домогосподарств, представлені неконтактні домогосподарства або ті, що тимчасово вибули з обстеження чи відмовились брати участь у ньому, то така страта вважається "неповною". Якщо ж таких домогосподарств немає, а є лише звичайні, то страта вважається "повною".

Для саморепрезентативних домогосподарств, тобто домогосподарств, площа землі яких є більшою за поріг саморепрезентативності, імовірність відбору в межах ПТОВ дорівнює

одиниці, $P_{2i} = 1$.

Для домогосподарств інших трьох страт імовірність відбору розраховується за формулами (3.7 – 3.9).

“Повна” ПТОВ:

У цьому випадку для всіх домогосподарств, розподілених по стратах, імовірність відбору визначається за формулою:

$$P_{2i} = n_{r,l,c} \cdot \frac{S_{r,l,c,i}}{S_{r,l,c}}, \quad (3.7)$$

де c – номер страти ($c = 1, 2, 3, 4$); $n_{r,l,c}$ – кількість усіх домогосподарств страти c , $S_{r,l,c}$ – їх загальна площа землі, $S_{r,l,c,i}$ – площа землі i -го домогосподарства c -ої страти l -ої ПТОВ r -го району.

“Неповна” ПТОВ:

У цьому випадку існують два варіанти:

а) звичайні домогосподарства відсутні у I, II або III страті.

У цьому випадку розрахунок імовірності відбору домогосподарств виконується за формулою, яка не враховує їх стратифікацію, але враховує наявність саморепрезентативних домогосподарств у ПТОВ:

$$P_{2i} = (n_{r,l} - n_{r,l}^*) \cdot \frac{S_{r,l,i}}{S_{r,l} - S_{r,l}^*}, \quad (3.8)$$

де $n_{r,l}$ – загальна кількість домогосподарств у ПТОВ, $S_{r,l,i}$ – площа землі i -го домогосподарства, $n_{r,l}^*$ – загальна кількість саморепрезентативних домогосподарств у ПТОВ, $S_{r,l}^*$ – їх загальна площа землі.

б) звичайні домогосподарства відсутні у IV страті (саморепрезентативні домогосподарства).

У цьому випадку розрахунок імовірності відбору домогосподарств виконується за формулою, яка не враховує їх стратифікацію та наявність саморепрезентативних домогосподарств у ПТОВ:

$$P_{2i} = n_{r,l} \cdot \frac{S_{r,l,i}}{S_{r,l}}, \quad (3.9)$$

де $n_{r,l}$ – загальна кількість домогосподарств у ПТОВ, $S_{r,l,i}$ – площа землі i -го домогосподарства, $S_{r,l}$ – площа землі l -ої ПТОВ r -го району.

3.1.3. Визначення загальної ймовірності відбору домогосподарств P_i та базової ваги W_{li}

Загальна ймовірність відбору домогосподарств розраховується окремо для кожного домогосподарства по кожній ПТОВ з урахуванням усіх можливих умов за формулою:

$$P_i = P_{li} \cdot P_{2i}. \quad (3.10)$$

Базовою вагою i -го домогосподарства є обернена величина загальної ймовірності відбору домогосподарства:

$$W_{li} = \frac{1}{P_i} = \frac{1}{P_{li}} \cdot \frac{1}{P_{2i}}. \quad (3.11)$$

3.2. Коригування базової ваги домогосподарств для врахування ступеня участі відібраних одиниць в обстеженні

Згідно з методологією обстеження СГД заміна відібраних домогосподарств, які не беруть участі в обстеженні внаслідок неможливості встановлення контакту з домогосподарством, або тимчасового його вибуття з обстеження, або відмови брати у ньому участь, не здійснюється. З метою компенсації інформації, недоотриманої внаслідок таких випадків, тобто наявності “неповних” страт, для підвищення надійності оцінювання показників за результатами обстеження здійснюється коригування базових ваг звичайних домогосподарств. Значення ваг зачинених домогосподарств залишаються без коригування, а ваги домогосподарств, які не беруть участі в обстеженні, приймаються рівними нулю.

Залежно від типу ПТОВ (“повна” чи “неповна”) використовують три коригуючі коефіцієнти (3.12, 3.14, 3.16).

“Повна” ПТОВ:

У цьому випадку коригуючий коефіцієнт розраховується для кожної страти за формулою:

$$k_{r,l,c} = \frac{n_{r,l,c} - n_{r,l,c}^{(3Ч)} - n_{r,l,c}^{(ББ)}}{n_{r,l,c}^{(3Б)}}, \quad (3.12)$$

де c – номер страти ($c = 1, 2, 3, 4$), $n_{r,l,c}$ – кількість усіх домогосподарств страти c , $n_{r,l,c}^{(3Ч)}$ – кількість звичайних домогосподарств страти c , $n_{r,l,c}^{(3Ч)}$ – кількість зачинених домогосподарств

страти c , $n_{r,l,c}^{(ББ)}$ – кількість домогосподарств страти c , які вибули з обстеження в середині року через переїзд або смерть господаря.

Тоді вага i -го домогосподарства розраховується за формулою:

$$W_{2i} = W_{li} \cdot k_{r,l,c}. \quad (3.13)$$

“Неповна” ПТОВ:

У цьому випадку існують два варіанти:

а) звичайні домогосподарства відсутні у I, II або III страті.

При визначенні коефіцієнта коригування не враховується стратифікація домогосподарств, а враховується лише наявність саморепрезентативних домогосподарств. В цьому випадку розраховується загальний коригуючий коефіцієнт для I, II та III страти за формулою:

$$k_{r,l,I-III} = \frac{n_{r,l,I-III} - n_{r,l,I-III}^{(3Ч)} - n_{r,l,I-III}^{(ББ)}}{n_{r,l,I-III}^{(3Б)}}, \quad (3.14)$$

де $n_{r,l,I-III}$ – кількість усіх домогосподарств за стратами I - III;

$n_{r,j,i-III}^{(IB)}$ – кількість звичайних домогосподарств за stratami I - III;

$n_{r,j,i-III}^{(3Y)}$ – кількість зачинених домогосподарств за stratami I - III;

$n_{r,j,i-III}^{(BB)}$ – кількість домогосподарств straty, які вибули з обстеження в середині року через переїзд або смерть господаря за stratami I - III.

Вага i -го домогосподарства розраховується за формулою:

$$w_{2i} = w_{1i} \cdot k_{r,j,i-III}. \tag{3.15}$$

б) звичайні домогосподарства відсутні у IV stratі (саморепрезентативні домогосподарства).

У цьому випадку коефіцієнт коригування визначається в цілому для ПТОВ за формулою:

$$k_{r,j,i-IV} = \frac{n_{r,j,i-IV} - n_{r,j,i-IV}^{(3Y)} - n_{r,j,i-IV}^{(BB)}}{n_{r,j,i-IV}^{(3B)}}, \tag{3.16}$$

де $n_{r,j,i-IV}$ – загальна кількість домогосподарств у ПТОВ; $n_{r,j,i-IV}^{(IB)}$ – загальна кількість звичайних домогосподарств у ПТОВ; $n_{r,j,i-IV}^{(3Y)}$ – загальна кількість зачинених домогосподарств у ПТОВ; $n_{r,j,i-IV}^{(BB)}$ – загальна кількість домогосподарств у ПТОВ, які вибули з обстеження в середині року через переїзд або смерть господаря.

Вага i -го домогосподарства розраховується за формулою:

$$w_{2i} = w_{1i} \cdot k_{r,j,i-IV}. \tag{3.17}$$

Якщо площі землі домогосподарств, що відмовилися, значно варіюють, коригуючі коефіцієнти (3.12, 3.14, 3.16) більш доцільно розраховувати на базі площі домогосподарств, а не на їх кількості.

3.3. Калібрація системи статистичних ваг

Для узгодження результатів, отриманих за даними обстеження СГД, з наявною зовнішньою інформацією застосовується процедура калібрації системи статистичних ваг.

Для обстеження СГД були визначені такі основні дані, отримані з зовнішніх джерел:

– загальна кількість сільських домогосподарств;

– загальна площа землі, що знаходиться у володінні та користуванні сільських домогосподарств.

Зовнішня інформація, яка використовувалась в обстеженні СГД для калібрації системи ваг у 2010 році, наведена у табл. 3.1.

Таблиця 3.1

Зовнішня інформація для калібрації системи статистичних ваг обстеження СГД у 2010 році

Регіон	Площа землі домогосподарств, га	Кількість домогосподарств
УКРАЇНА	7 301 990,3	5 304 794
Автономна Республіка Крим	326 125,4	224 625
Вінницька	309 594,6	347 712
Волинська	311 761,0	163 738
Дніпропетровська	391 039,7	215 166
Донецька	306 599,3	170 305
Житомирська	234 994,5	208 930
Закарпатська	231 720,1	236 477
Запорізька	407 655,2	160 986
Івано-Франківська	262 487,8	269 865
Київська	209 970,6	299 316
Кіровоградська	298 337,2	153 315
Луганська	185 671,1	115 734
Львівська	329 299,4	316 175
Миколаївська	362 780,6	132 411
Одеська	451 856,3	290 884
Полтавська	307 202,0	236 200
Рівненська	265 643,3	191 092
Сумська	157 413,6	152 134
Тернопільська	263 009,4	202 539
Харківська	321 388,4	211 575
Херсонська	551 432,1	150 557
Хмельницька	226 363,8	238 627
Черкаська	200 257,7	246 158
Чернівецька	208 087,6	187 705
Чернігівська	181 299,6	182 568

Масив зовнішньої інформації для калібрації формується один раз на рік у період обробки даних базового інтерв'ю та використовується без змін при калібрації результатів усіх щомісячних інтерв'ю обстеження СГД поточного року.

Побудова системи рівнянь для калібрації ваг для обстеження СГД у травні 2008 – квітні 2013 року, що задовольняє наведену в таблиці 3.1 структуру зовнішніх даних, у межах окремого регіону має вигляд:

$$\begin{cases} \sum_{i=1}^n \frac{1}{2} \cdot \frac{(w_i^{(c)} - w_i)^2}{w_i} \rightarrow \min; \\ \sum_{i=1}^n w_i^{(c)} = N; \\ \sum_{i=1}^n s_{r,j,i} \cdot w_i^{(c)} = S; \end{cases} \tag{3.18}$$

де $w_i, i = 1, \dots, n$ – система статистичних ваг до калібрації, $w_i^{(c)}, i = 1, \dots, n$ – система статистичних ваг після калібрації, $s_{r,j,i}$ – площа землі i -го домогосподарства, N – кількість сільських домогосподарств за зовнішніми даними регіону, S – загальна площа земельних ділянок сільських домогосподарств у регіоні.

Розв'язання задачі (3.18) здійснюється стандартними математичними методами, зокрема методом множників Лагранжа за допомогою сучасних пакетів програм для реалізації процедур калібрації системи статистичних ваг. Одним із таких пакетів є g-Calib, який розробляє офіс статистики Бельгії для застосування в середовищі SPSS [9].

На основі процедури калібрації визначаються вагові коефіцієнти $k_i^{(c)}$, що забезпечують відповідність результатів обстеження встановленим обмеженням та мінімізують відхилення ваг до та після калібрації.

Після визначення коефіцієнтів коригування ваг $k_i^{(c)}$ розраховуються калібраційні ваги домогосподарств w_{3i} за формулою:

$$w_{3i} = w_{2i} \cdot k_i^{(c)}. \tag{3.19}$$

Зазначимо, що за результатом застосування окремих процедур калібрації може розраховуватись безпосередньо система калібраційних ваг w_{3i} . За таких умов коефіцієнти коригування ваг $k_i^{(c)}$ доцільно розрахувати окремо за формулою:

$$k_i^{(c)} = \frac{w_{3i}}{w_{2i}}. \tag{3.20}$$

Це необхідно для забезпечення однорідності процедур контролю за системою ваг.

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

3.4. Приклад розрахунку статистичних ваг

У якості прикладу розглянемо розрахунок імовірності включення до вибірки ПТОВ і домогосподарств Бердянського району Запорізької області, який відноситься до однієї ПСТ.

За результатами відбору до вибірки включено п'ять ПТОВ, створених із сільських рад, домогосподарства яких представляють усі сільські домогосподарства району (табл.3.2., рис.3.1.).

Загальна площа землі домогосподарств району S_p становить 22219,70 га. За формулою (3.1) порогове значення площі землі домогосподарств у районі дорівнює 4443,94 га, що більше, ніж площа землі кожної відібраної ПТОВ. Отже, самореєзентативних ПТОВ у районі немає, а імовірності відбору ПТОВ розраховуються за формулою (3.3) і наводяться у таблиці 3.2.

Таблиця 3.2.

Результати відбору ПТОВ Бердянського району Запорізької області для обстеження СГД у травні 2008 – квітні 2013 року

Коди КОАТУУ			Назва ПТОВ	Площа землі домогосподарств, га	Імовірність відбору ПТОВ, P_{Li}
регіону	району	місцевої ради			
23	206	815	Дмитрівська	3620,2	0,8146
23	206	850	Новотроїцька	2016,4	0,4537
23	206	860	Червонопільська	1998,8	0,4498
23	206	805	Андрівська	1798,2	0,4046
23	206	845	Новопетрівська	520,8	0,1172

Величина P_{Li} для кожної ПТОВ характеризує частку відрізка загальною довжиною 4443,94 га, яка відповідає площі землі цієї ПТОВ (рис.3.1).

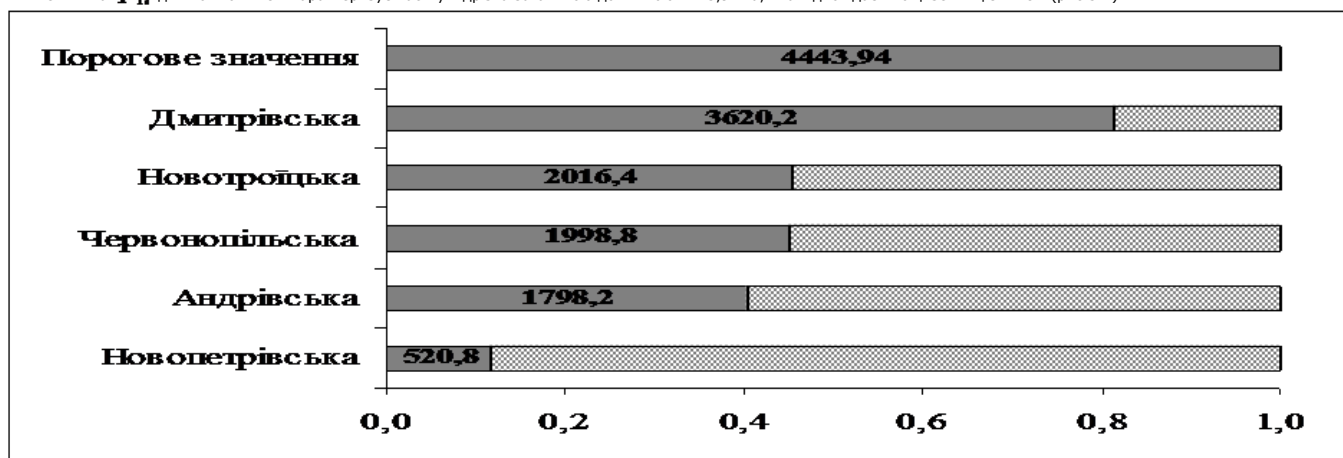


Рис.3.1. Імовірності включення до вибірки ПТОВ Бердянського району Запорізької області для обстеження СГД у травні 2008 – квітні 2013 року

Розглянемо розрахунок імовірностей відбору, базових ваг, коригуючих коефіцієнтів і скоригованих базових ваг домогосподарств для Андрівської та Червонопільської ПТОВ (табл.3.3).

За результатами відбору в Андрівській ПТОВ із загальною площею землі домогосподарств 1798,2 га відібрано 13 домогосподарств. Згідно з розрахунками за формулою (3.5) поріг самореєзентативності $S_{np}^{(II)}$ дорівнює 138,32 га. Домогосподарству під кодом 102, у якого площа землі більша за цей поріг, присвоюється значення імовірності відбору P_{21} , що дорівнює

1. За формулою (3.6) $S_{np}^{*(II)}$ дорівнює 125,79 га. Як видно з таблиці 3.3, площі землі домогосподарств, які залишилися, менші за це значення, тобто всі ці домогосподарства є несамореєзентативними. Оскільки кожна страта Андрівської ПТОВ представлена звичайними домогосподарствами, то імовірність відбору домогосподарств цієї ПТОВ розраховується за формулою (3.7), а коригуючий коефіцієнт – за формулою (3.12).

У Червонопільській ПТОВ, яка має загальну площу землі 1998,8 га, відібрано 14 домогосподарств. Поріг самореєзентативності $S_{np}^{(II)}$ дорівнює 142,77 га, а $S_{np}^{*(II)}$ дорівнює

104,35 га. Тому єдиним самореєзентативним домогосподарством є домогосподарство з площею землі 642,19 га під кодом 211, імовірність відбору якого P_{21} дорівнює 1. Оскільки воно відмовилось брати участь у обстеженні, то ця ПТОВ є "неповною" і імовірність відбору інших домогосподарств розраховується за формулою (3.9), а коригуючий коефіцієнт – за формулою (3.16).

Таблиця 3.3.

Результати відбору домогосподарств Андрівської та Червонопільської ПТОВ Бердянського району Запорізької області для обстеження СГД у вересні 2010 року

№	Коди КОАТУУ			Код домогосподарства	Площа землі домогосподарства, га	Страта	Ознака участі	Імовірність відбору ПТОВ, P_{Li}	Імовірність відбору домогосподарства,	Базова вага, w_{Li}	Коригуючий коефіцієнт, k	Базова вага, скоригована на	Калібраційний коефіцієнт коригування ваг,	Калібраційна вага, w_{3i}
	регіону	району	місцевої ради											

										P_{2j}			коефіцієнт,	$k_L^{(j)}$		
1	23	206	805	102	288,77	IV	3В	0,4046	1,0000	2,4716	-	2,4716	1,8249	4,5104		
2	23	206	805	111	25,91	III	3В	0,4046	0,0456	54,2012	2,0	108,4024	0,9619	104,2723		
3	23	206	805	101	2,00	III	ВМ	0,4046	0,0035	706,1648	-	0,0000	-	-		
4	23	206	805	109	1,00	II	3В	0,4046	0,0313	78,9641	-	78,9641	0,8800	69,4884		
5	23	206	805	112	1,00	II	3В	0,4046	0,0313	78,9641	-	78,9641	0,8800	69,4884		
6	23	206	805	113	0,93	II	3В	0,4046	0,0291	84,9339	-	84,9339	0,8797	74,7164		
7	23	206	805	110	0,83	II	3В	0,4046	0,0260	95,0606	-	95,0606	0,8795	83,6058		
8	23	206	805	106	0,61	II	3В	0,4046	0,0191	129,4019	-	129,4019	0,8787	113,7054		
9	23	206	805	107	0,54	II	3Ч	0,4046	0,0169	146,2472	-	146,2472	0,8748	127,9371		
10	23	206	805	105	0,48	I	3В	0,4046	0,0106	233,1676	-	233,1676	0,8783	204,7911		
11	23	206	805	103	0,37	I	3В	0,4046	0,0082	301,4118	-	301,4118	0,8779	264,6094		
12	23	206	805	104	0,28	I	3В	0,4046	0,0062	398,6414	-	398,6414	0,8776	349,8477		
13	23	206	805	108	0,21	I	3В	0,4046	0,0046	537,2993	-	537,2993	0,8774	471,4264		
14	23	206	860	211	642,19	IV	ВМ	0,4498	1,0000	2,2232	-	0,0000	-	-		
15	23	206	860	204	44,70	III	3В	0,4498	0,3131	7,1006	1,0769	7,6466	1,0239	7,8294		
16	23	206	860	213	3,50	III	3В	0,4498	0,0245	90,7433	1,0769	97,7215	0,8882	86,7962		
17	23	206	860	201	1,00	II	3В	0,4498	0,0070	317,6015	1,0769	342,0251	0,8800	300,9821		
18	23	206	860	202	0,90	II	3В	0,4498	0,0063	352,8905	1,0769	380,0278	0,8797	334,3105		
19	23	206	860	207	0,80	II	3В	0,4498	0,0056	397,0018	1,0769	427,5312	0,8793	375,9282		
20	23	206	860	212	0,72	II	3В	0,4498	0,0050	444,6421	1,0769	478,8351	0,8791	420,9439		
21	23	206	860	209	0,65	II	3В	0,4498	0,0046	483,3066	1,0769	520,4729	0,8789	457,4436		
22	23	206	860	210	0,58	II	3В	0,4498	0,0041	542,2464	1,0769	583,9451	0,8786	513,0542		
23	23	206	860	206	0,50	I	3В	0,4498	0,0035	635,2029	1,0769	684,0500	0,8784	600,8695		
24	23	206	860	205	0,47	I	3В	0,4498	0,0033	673,7001	1,0769	725,5076	0,8783	637,2133		
25	23	206	860	214	0,44	I	3В	0,4498	0,0031	717,1646	1,0769	772,3146	0,8782	678,2467		
26	23	206	860	208	0,38	I	3В	0,4498	0,0027	823,4112	1,0769	886,7315	0,8780	778,5503		
27	23	206	860	203	0,28	I	3В	0,4498	0,0020	1111,6052	1,0769	1197,0876	0,8776	1050,5641		

* 3В – звичайне домогосподарство; ВМ – домогосподарство, яке відмовилось від обстеження; 3Ч – зачинене домогосподарство.

Порядок визначення якості системи ваг обстеження сільськогосподарської діяльності населення в сільській місцевості

Якість системи ваг обстеження СГД визначається після кожного етапу їх коригування, а саме після визначення загальної ймовірності відбору домогосподарств та базової ваги, коригування ваг для врахування ступеня участі відібраних одиниць в обстеженні та калібрації системи ваг (див. п. 3.1 – 3.3) окремо за результатами основного та всіх щомісячних інтерв'ю.

Розрахунки здійснюються на національному рівні, тобто по всьому агрегованому масиву даних обстеження.

При контролі якості системи ваг звичайно на практиці розраховуються такі характеристики:

1. Середня вага по всіх домогосподарствах як окремий множник – коефіцієнт інфляції $F_{(j)}$. Величина $F_{(j)}$ для j -го етапу розрахунку ваг визначається за формулою:

$$F_{(j)} = \frac{\sum_{i=1}^n w_{(j)i}}{n}. \quad (4.1)$$

Коефіцієнт інфляції є важливим параметром системи ваг, який показує:

- скільки в середньому домогосподарств репрезентує одне обстежене домогосподарство на певному етапі обстеження СГД;
- яку площу землі домогосподарств генеральної сукупності представляє в середньому один гектар площі обстеженого домогосподарства.

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

При аналізі отриманих результатів необхідно звернути особливу увагу на те, що $w_{\min} \in 0$ є неприпустимим. За таких умов, як правило, необхідно перевірити процедури розрахунку системи ваг і в разі необхідності забезпечити додатні значення величин ваг із застосуванням спеціальних процедур [9].

3. З огляду на незначну кількість випадків від'ємних ваг і незначний їх вплив на кінцеві результати оцінювання показників обстеження СГД, значення ваги w_i приймають за одиницю, тобто $w_i = 1$.

3. Розраховуються також співвідношення між коефіцієнтом інфляції та мінімальним і максимальним значеннями ваг – $R_{(j)1}$, $R_{(j)2}$ та $R_{(j)3}$ – за формулами:

$$R_{(j)1} = \frac{F_{(j)}}{w_{(j)\min}}; R_{(j)2} = \frac{w_{(j)\max}}{F_{(j)}}; R_{(j)3} = \frac{w_{(j)\max}}{w_{(j)\min}}. \quad (4.2)$$

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.

4. Розмах варіації ваг, як характеристики їх варіативності, розраховується за формулою:

$$\Delta R_{(j)} = w_{(j)\max} - w_{(j)\min}. \quad (4.3)$$

5. Відносне зростання дисперсії вибірових оцінок, джерелом яких є варіативність ваг одиниць вибірки, k_L . Величина k_L визначається за формулою [4]:

$$k_L = \frac{n \sum_{i=1}^n (w_{(j)i})^2}{(\sum_{i=1}^n w_{(j)i})^2}. \quad (4.4)$$

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

6. Коефіцієнт варіації ваг $CV(w_{(j)})$ розраховується за формулою:

$$CV(w_{(j)}) = \sqrt{\frac{1}{n \cdot F_{(j)}^2} \cdot \sum_{i=1}^n (w_{(j)i} - F_{(j)})^2}; \quad (4.5)$$

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

7. Відношення коефіцієнтів інфляції для поточного та попереднього етапів зважування на j -му етапі $R_{F(j)}$ визначається за формулою:

$$R_{F(j)} = \frac{F_{(j)}}{F_{(j-1)}}. \quad (4.6)$$

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

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

$$R_{(j)4} = \frac{CV(w_{(j)})}{CV(w_{(j-1)})}. \quad (4.7)$$

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

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

$$r_{(j,j-1)} = \frac{\sum_{i=1}^n (w_{(j-1)i} - \bar{w}_{(j-1)}) \cdot (w_{(j)i} - \bar{w}_{(j)})}{\sqrt{\sum_{i=1}^n (w_{(j-1)i} - \bar{w}_{(j-1)})^2 \cdot \sum_{i=1}^n (w_{(j)i} - \bar{w}_{(j)})^2}}. \quad (4.8)$$

Коефіцієнт кореляції (4.8) характеризує щільність лінійного статистичного зв'язку систем ваг, що відповідають суміжним етапам розрахунку ваг. Чим ближчим є модуль величини цього коефіцієнта до одиниці, тим менше розподіл поточних ваг відхиляється від розподілу попереднього етапу й, у цілому від початкових ваг дизайну.

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

$$N_{(j)} = \sum_{i=1}^n w_{(j)i}, \quad (4.9)$$

Оцінка загальної кількості домогосподарств $N_{(j)}$ порівнюється з оцінкою загальної кількості домогосподарств із зовнішніх джерел N_{3l} . Розраховується абсолютне відхилення цих оцінок за формулою:

$$D_{(j)} = N_{3l} - N_{(j)}. \quad (4.10)$$

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

У таблиці 4.1 наведено приклад результатів оцінювання якості кінцевої системи ваг обстеження СГД за травень 2006 - 2010 років, який ілюструє зміну окремих характеристик якості ваг базового обстеження.

Таблиця 4.1.

Зміна окремих характеристик якості кінцевої системи ваг обстеження СГД за травень 2006 – 2010 років

Характеристики	Рік				
	2006	2007	2008	2009	2010
Середня вага, F	190,12	187,85	191,11	188,41	185,46
Мінімальна вага, w_{\min}	0,30	0,56	0,04	0,47	0,47
Максимальна вага, w_{\max}	20 967,42	43 490,63	46 345,59	28 320,53	42 066,55
Коефіцієнт варіації ваг, $CV(w)$, %	2,36	2,63	3,16	3,01	3,24
$R_1 = \frac{F}{w_{\min}}$	633,73	335,45	4 777,75	400,87	390,88
$R_2 = \frac{w_{\max}}{F}$	110,29	231,52	242,51	150,31	226,82
$R_3 = \frac{w_{\max}}{w_{\min}}$	69 891,40	77 661,84	1 158 639,75	60 256,45	89 503,30
Розмах варіації ваг, ΔR	20 967,11	43 490,07	46 345,55	28 320,06	42 066,07

Приклад зміни окремих характеристик якості базової та кінцевої системи ваг протягом п'яти циклів обстеження 2006-2007, 2007-2008, 2008-2009, 2009-2010 та 2010-2011 років (період із травня по вересень) наведено на рис.4.1 – 4.3.

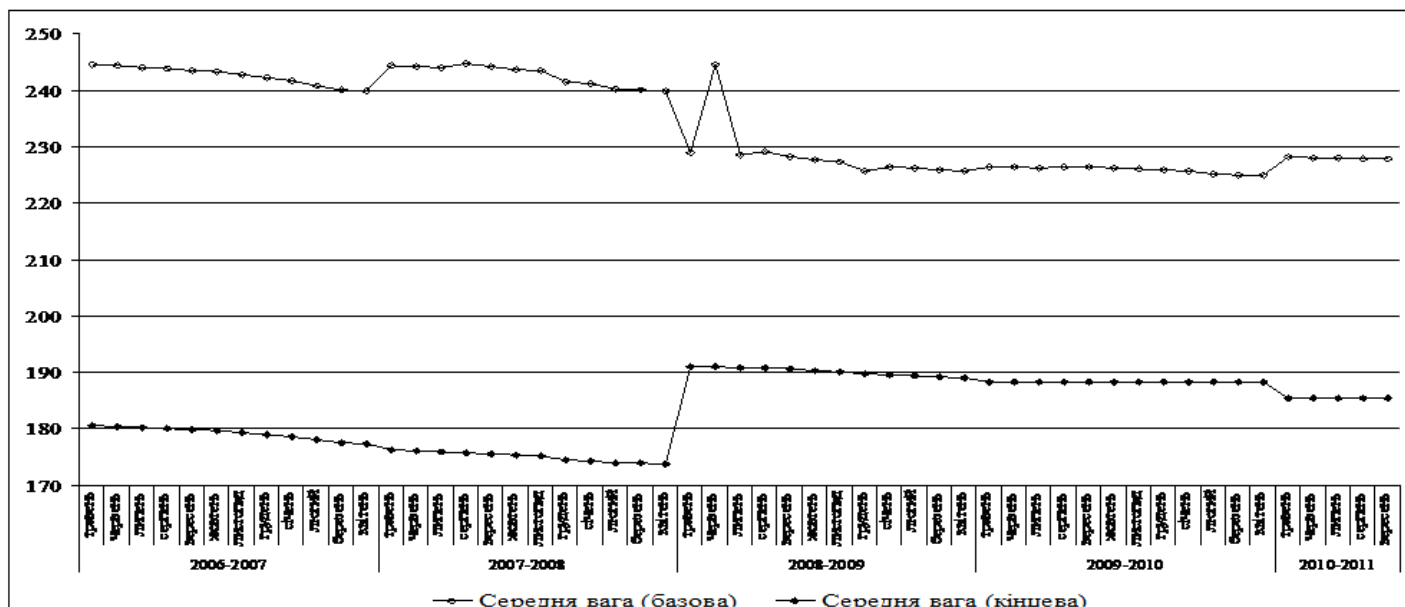


Рис.4.1. Середня вага по всіх домогосподарствах

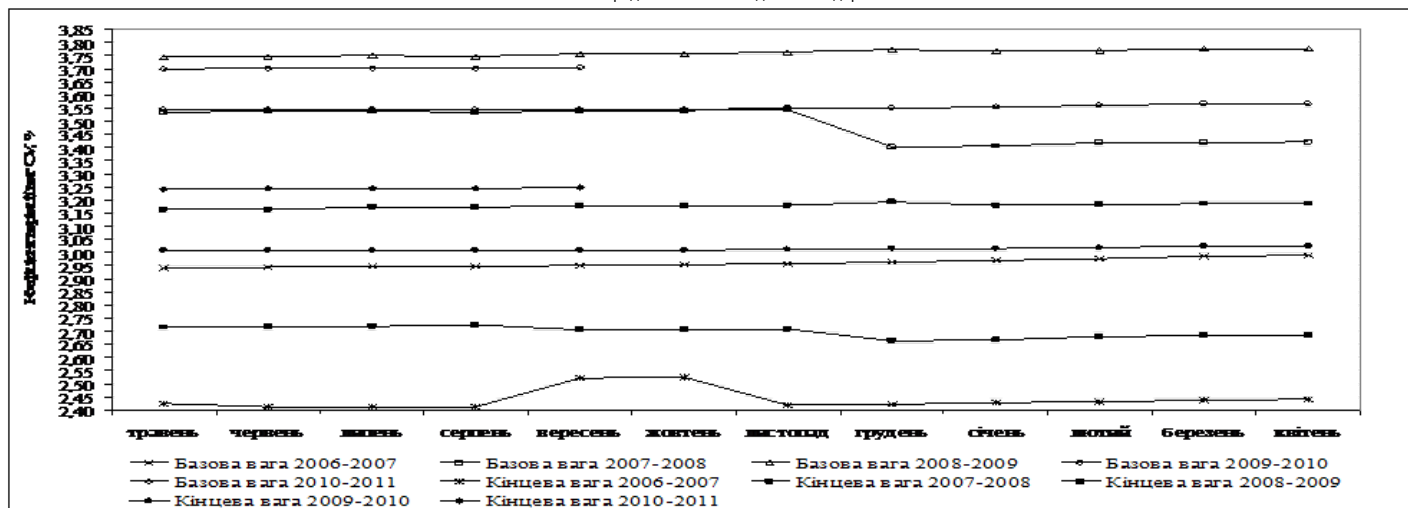


Рис.4.2 Коефіцієнт варіації ваг

Рис.4.3 Відносне зростання дисперсії вибірових оцінок

Як видно з наведеного, величина середньої ваги залишається майже незмінною протягом річного циклу обстеження і дещо зменшується з року в рік. Ураховуючи, що обсяг щорічної вибірки практично не змінювався, це можна пояснити головним чином зменшенням кількості домогосподарств у генеральній сукупності. Різка зміна середньої ваги у травні 2008 року пов'язана зі зміною вибіркової сукупності ПТОВ.

Величина відносного зростання вибірових оцінок є значною і відносно стабільною. Причому для кінцевих ваг вона менша, ніж для базових ваг. Це пояснюється впливом процедури калібрації ваг, яка покращує їх якість і зменшує варіативність.

ДЖЕРЕЛА

1. Deville J.-C., Sarndal C.-E. Calibration Estimators in Survey Sampling // Journal of the American Statistical Association. – 1992. – Vol. 87, № 418.
2. Deville J.-C., Sarndal C.-E., Sautory O. Generalized Raking Procedures in Survey Sampling // Journal of the American Statistical Association. – 1993. – Vol. 88, № 423.
3. Kalton G. Introduction to Survey Sampling. – Beverly Hills: Sage, 1983.
4. Kish L.: Selected papers. – New York: Wiley, 2003.
5. Kish L. Survey sampling. – New York: Wiley, 1995.
6. Вибіркове спостереження: термінологічний словник / [наук.кер. Васечко О.О.]. – К.: НТК статистичних досліджень, 2004. – 140 с.
7. Методика формування вибірових сукупностей для проведення у 2009-2013 роках вибірових обстежень населення (домогосподарств): умов життя домогосподарств, економічної активності населення та сільськогосподарської діяльності населення у сільській місцевості, затверджена наказом Держкомстату від 14.08.09 № 308.
8. Наказ Державного комітету статистики України від 01.09.2000р. №290 "Про затвердження інструментарію вибірового обстеження сільського-подарської діяльності населення".
9. Саріогло В.Г. Проблеми статистичного зважування вибірових даних. – К.: ІВЦ Держкомстату України, 2005.