

Державна служба статистики України

ЗАТВЕРДЖЕНО

**Наказ Державної служби
статистики**

13 листопада 2023 року № 313

**МЕТОДИКА
РОЗРАХУНКУ СИСТЕМИ СТАТИСТИЧНИХ ВАГ ДЛЯ ПОШИРЕННЯ
РЕЗУЛЬТАТІВ ДЕРЖАВНОГО СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ
"ОБСТЕЖЕННЯ РОБОЧОЇ СИЛИ" НА ГЕНЕРАЛЬНУ СУКУПНІСТЬ**

Державна служба статистики України

Відповідальний за підготовку – директор департаменту обстежень домогосподарств апарату Держстату Алла Солоп.

Методика розрахунку системи статистичних ваг для поширення результатів державного статистичного спостереження "Обстеження робочої сили" на генеральну сукупність (далі – Методика) визначає основні принципи та порядок розрахунку системи статистичних ваг домогосподарств і осіб в обстеженні робочої сили.

Методика ґрунтується на сучасних методологічних засадах поширення на генеральну сукупність результатів вибірових обстежень населення, що використовуються в міжнародній практиці, і є однією з основних складових методичного забезпечення проведення обстеження робочої сили.

Ця Методика призначена для використання працівниками органів державної статистики, а також може бути корисною для інших заінтересованих користувачів статистичної інформації.

Методику підготували фахівці відділу обстеження робочої сили департаменту обстежень домогосподарств Держстату та фахівець Інституту демографії та соціальних досліджень ім. М. В. Птухи НАН України Володимир Саріогло – д.е.н., завідувач відділу моделювання соціально-економічних процесів і структур.

Методика схвалена Комісією з питань удосконалення методології та звітної документації Державної служби статистики України (протокол від 22 вересня 2023 року № КПУМ/22-23).

Держстат
адреса: вул. Шота Руставелі 3, Київ, 01601
<http://www.ukrstat.gov.ua>
телефон: (044) 287-24-22
адреса електронної пошти: office@ukrstat.gov.ua

Зміст

	Стор.
Скорочення та умовні позначення.....	4
I. Вступ.....	5
II. Джерела інформації для реалізації процедур розрахунку системи статистичних ваг.....	6
III. Принципи оцінювання показників.....	7
IV. Порядок розрахунку складових системи статистичних ваг.....	11
1. Розрахунок базової ваги домогосподарств.....	11
2. Урахування відмов домогосподарств від обстеження.....	11
3. Урахування відмов осіб від обстеження.....	12
4. Урахування статевовікової структури.....	13
V. Порядок визначення якості системи статистичних ваг.....	14
Додатки:	
Додаток 1. Схема визначення системи статистичних ваг для оцінки показників за результатами обстеження робочої сили.....	18
Додаток 2. Ваги домогосподарств і коефіцієнти коригування статистичних ваг для домогосподарств Любешівського району Волинської області за результатами обстеження робочої сили у 2021 році.....	19
Додаток 3. Окремі характеристики якості системи статистичних ваг обстеження робочої сили у 2019–2021 роках.....	20
Список використаних джерел.....	21

Скорочення та умовні позначення

ДСС,
статистичне
спостереження
ЄС
ПТОВ
ВТОВ

- державне статистичне спостереження;
- Європейський Союз;
- первинна територіальна одиниця вибірки;
- вторинна територіальна одиниця вибірки.

I. Вступ

1. Обстеження робочої сили (далі – ОРС) є державним статистичним спостереженням, результати якого призначені для інформування користувачів щодо обсягів та змін пропозиції робочої сили (зайнятості, безробіття) на ринку праці України, аналізу тенденцій розвитку його кон'юнктури.

2. Генеральну сукупність ОРС складають приватні домогосподарства та особи у віці 15 років і старше – члени цих домогосподарств, які проживають на території України. Особи, які проживають у колективних домогосподарствах і в інституційних закладах, не належать до генеральної сукупності ОРС.

3. Методика розроблена з метою визначення основних принципів поширення результатів ОРС на генеральну сукупність, порядку розрахунку складових системи статистичних ваг та формування кінцевих статистичних ваг осіб і домогосподарств.

4. Методика враховує положення Регламенту (ЄС) 2019/1700 від 10 жовтня 2019 року [9] та Виконавчого регламенту Комісії (ЄС) 2019/2400 від 16 грудня 2019 року [10].

5. Основним методологічним принципом, що визначає можливість поширення результатів на генеральну сукупність, є принцип імовірності (репрезентативності) вибірки. Цей принцип забезпечує наближеність оцінок показників, отриманих за результатами статистичного спостереження, до їх дійсних значень у генеральній сукупності. Зазначений принцип полягає в тому, що при формуванні вибірки кожен елемент генеральної сукупності має певну наперед визначену ненульову ймовірність потрапити до вибірки і ця ймовірність має бути врахована на етапі оцінювання показників.

6. Можливість дотримання принципу ймовірності в ОРС ускладнюється тим, що це ДСС має складний дизайн вибірки, характеризується значними й неоднорідними за регіонами та групами населення рівнями відмов від участі в статистичному спостереженні, вимагає застосування складних методів обробки даних, оцінювання показників і узгодження результатів обстеження з наявною зовнішньою інформацією. Відповідно для забезпечення репрезентативності результатів ДСС необхідною є реалізація спеціальної багатоступеневої процедури їх статистичного зважування. Це зумовлено тим, що саме 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 (далі – зовнішні джерела).

7. Система статистичних ваг є невід'ємною складовою результатів ОРС. Вона розраховується з метою визначення результуючої ваги кожної одиниці

спостереження, що була обстежена, та оцінювання показників за генеральною сукупністю за результатами обстеження.

8. Відповідно до методології ОРС статистичні ваги розраховуються для домогосподарств і осіб, які входять до їх складу.

9.3 використанням системи статистичних ваг також реалізуються процедури поліпшення якості оцінювання показників на основі складних оціночних функцій (оцінки за відношенням, оцінки за регресією тощо) та узгодження результатів обстеження з інформацією зовнішніх джерел (зокрема із застосуванням процедур калібрації). Вона дозволяє об'єднувати результати різних вибірових обстежень на нижчих рівнях агрегації даних.

10. Для цілей цієї Методики терміни вживаються у значеннях, наведених у статтях 1, 8, 9, 10 Закону України "Про офіційну статистику" [1], у пунктах 2, 9, 13, 15, 54 Глосарія до плану статистичного спостереження [3], у пунктах A006, B012, B020, B054, G008, D001, D010, D014, D015, D020, D022, Z012, K001, K006, K022, K032, O011, O029, P007, P061, P084, P092, R002, R012, C033, C043, C053, C060, F013 Глосарія термінів національної моделі діяльності органів державної статистики [4], Методиці формування вибірових сукупностей для проведення у 2019–2024 роках вибірових обстежень населення (домогосподарств): умов життя домогосподарств, робочої сили та сільськогосподарської діяльності населення у сільській місцевості [6], Методологічних положеннях щодо формування вибірових сукупностей для проведення органами державної статистики України вибірових обстежень населення (домогосподарств) [5], термінологічному словнику щодо вибірових обстежень [2], статистичному словнику [8].

II. Джерела інформації для реалізації процедур розрахунку системи статистичних ваг

1. Джерелами інформації для реалізації процедур розрахунку системи статистичних ваг є:

дані щодо дизайну та основних параметрів вибірки (характеристики вибіркової сукупності);

дані, отримані за результатами ОРС;

дані демографічної статистики.

2. До даних щодо дизайну й основних початкових параметрів вибірки належать:

загальнодержавний номер територіальної одиниці вибірки – ПТОВ, ВТОВ;
імовірність відбору ПТОВ;

імовірність відбору домогосподарств у межах ВТОВ у міській місцевості;

імовірність відбору ВТОВ у міській або домогосподарств у сільській місцевості;

загальна кількість домогосподарств у ВТОВ у міській місцевості;
загальна кількість домогосподарств у ПТОВ у сільській місцевості;
фактична кількість відібраних для обстеження домогосподарств у територіальній одиниці вибірки.

3. Початкові параметри вибірки залишаються незмінними протягом усього терміну дії територіальної вибірки для ОРС, яка триває п'ять років. Разом з тим у випадку вимушеної заміни територіальних одиниць вибірки або зміни фактичної кількості відібраних домогосподарств необхідно скоригувати відповідні параметри вибірки.

4. Дані для розрахунку системи статистичних ваг, отримані за результатами ОРС, включають: код домогосподарства, код особи, стать особи, вік особи, статус участі в робочій силі, загальнодержавний код територіальної одиниці, код регіону, тип місцевості, ознаку участі відібраних одиниць у спостереженні.

5. Дані демографічної статистики включають інформацію щодо чисельності населення за п'ятирічними віковими групами за статтю та типом місцевості в розрізі регіонів на 1 січня поточного року.

III. Принципи оцінювання показників

1. ОРС є вибіркоvim обстеженням складного дизайну, тому для адекватного й повного відображення явищ, 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, зокрема й ОРС.

2. Важливість цього етапу зумовлена тим, що некоректна система статистичних 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 імовірностей відбору одиниць вибірки, фактичних рівнів відмов домогосподарств і окремих осіб від участі в обстеженні, гармонізації результатів обстеження з надійними даними із зовнішніх джерел, зокрема з даними демографічної статистики щодо кількості та статевовікової структури населення. Отже, система статистичних ваг відображає дизайн вибірки, а також ураховує основні етапи реалізації обстеження.

3. Порядок формування вибіркової сукупності для проведення ДСС відповідно один раз на п'ять років викладений у Методиці формування вибірових сукупностей для проведення у 2019–2024 роках вибірових

обстежень населення (домогосподарств): умов життя домогосподарств, робочої сили та сільськогосподарської діяльності населення у сільській місцевості [6].

4. Розрахунок системи статистичних ваг складається з таких основних етапів (додаток 1 до цієї Методики):

- розрахунок базових ваг домогосподарств;
- коригування ваг для врахування ступеня участі домогосподарств в обстеженні;
- коригування ваг для врахування ступеня участі осіб в обстеженні;
- коригування ваг для узгодження результатів обстеження з наявною інформацією із зовнішніх джерел та визначення кінцевої ваги кожного обстеженого домогосподарства та особи;
- контроль якості системи статистичних ваг.

5. У загальному випадку кінцева вага r -ї особи, яка є членом i -го домогосподарства, розраховується як добуток базової ваги домогосподарства та відповідних вагових коефіцієнтів (коефіцієнтів переважування) за формулою:

$$w_{i,r} = w_{Bi} \times k_{1i} \times k_{2r} \times k_{3r}, \quad (1)$$

де $w_{i,r}$ – кінцева вага особи, яка є членом домогосподарства;

w_{Bi} – базова вага домогосподарства;

k_{1i}, k_{2r}, k_{3r} – вагові коефіцієнти.

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

6. Величина базової ваги є результатом певної послідовності розрахунків. Реалізований в обстеженні дизайн вибірки зумовлює використання для розрахунку базових ваг по кожній страті таких складових:

- імовірність відбору ПТОВ – P_1 ;
- імовірність відбору ВТОВ у міській місцевості або домогосподарств у сільській місцевості – P_2 ;
- імовірність відбору домогосподарств у межах ВТОВ у міській місцевості – P_3 .

Імовірності P_1, P_2, P_3 визначаються за результатами формування вибірки ОРС.

7. У ймовірнісній вибірці ОРС кожне i -те домогосподарство генеральної сукупності обсягу N , $i = 1, 2, \dots, N$ характеризується певною позитивною ненульовою ймовірністю p_i потрапити до вибірки. Тобто це домогосподарство репрезентує $1/p_i$ домогосподарств генеральної сукупності й відповідно при розрахунках для генеральної сукупності його потрібно врахувати $1/p_i$ разів. Величина $1/p_i$ є базовою статистичною вагою i -го домогосподарства.

8. Базові ваги розраховуються при формуванні вибірки на весь термін її експлуатації.

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

10. Теоретично обґрунтованим методом урахування випадків неодержання 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.

11. Коригування статистичних ваг для компенсації відмов респондентів від участі в обстеженні включає такі кроки:

розділення вибірки на систему комірок – класів зважування, що охоплюють усю вибірову сукупність і не перетинаються;

розрахунок єдиних коефіцієнтів коригування статистичних ваг для всіх одиниць певної комірки, як величини оберненої до частки одержаних відповідей від одиниць спостереження (розрахунки здійснюються окремо для домогосподарств і осіб);

контроль коефіцієнтів коригування статистичних ваг із метою запобігання дуже значним коригуванням статистичних ваг;

коригування базових статистичних ваг домогосподарств і осіб на основі визначених коефіцієнтів.

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

13. Доцільність узгодження результатів обстеження з інформацією щодо певних характеристик генеральної сукупності із зовнішніх джерел зумовлена такими обставинами:

оцінки показників за результатами вибірового обстеження завжди характеризуються певною похибкою, джерелами якої є несущільність обстеження (похибка вибірки) і наявність так званих невибірових похибок;

вибіркове обстеження населення не може забезпечити оцінок певних характеристик генеральної сукупності, наприклад оцінки загальної кількості населення або загальної кількості домогосподарств. Хоча ці оцінки й отримуються за результатами обстеження, вони відображають головним чином параметри, що були закладені при формуванні вибірки;

основні характеристики генеральної сукупності змінюються в часі, і відповідно окремі з них на момент завершення обробки даних обстеження дещо відрізняються від таких, якими вони були на початку п'ятирічного циклу дії територіальної вибірки в обстеженні.

14. З відомих на сьогодні узагальнених методів коригування системи статистичних ваг найбільш теоретично розробленою й ефективною є так звана процедура калібрації статистичних ваг, сутність якої полягає в трансформації статистичних 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.

15. Метод постстратифікації системи статистичних ваг є спрощеним варіантом методу калібрації та застосовується у випадку, коли наявна інформація зовнішніх джерел по певних стратах генеральної сукупності, які охоплюють усю сукупність і не перетинаються. Зазначені страти, як правило, відрізняються від страт, що були побудовані при формуванні вибіркової сукупності. За таких умов коефіцієнти для коригування системи статистичних ваг визначаються по кожній страті окремо як відношення значення показника за даними із зовнішніх джерел до оцінки показника за результатами обстеження або як коефіцієнти, що приводять структуру генеральної сукупності, оцінену за результатами обстеження, до її структури, визначеної за інформацією із зовнішніх джерел.

16. Для постстратифікації статистичних ваг в ОРС використовуються насамперед дані статистики населення щодо кількості, розміщення та статевовікової структури населення (далі – дані демографічної статистики).

17. На практиці вибірка завжди відхиляється від характеристик, передбачених її дизайном, за рахунок таких факторів, як наявність у списках для відбору адрес неіснуючих домогосподарств, відхилення навантаження фахівців з інтерв'ювання з обстежень від визначеної величини, відмов домогосподарств та їх членів від обстеження, що мають різний рівень у різних територіальних одиницях, тощо. Кінцева система статистичних ваг одиниць спостереження має враховувати такі відхилення для підвищення надійності оцінювання показників за результатами обстеження. Ураховуючи необхідність реалізації процедур коригування статистичних ваг та їх вплив на якість оцінок показників за результатами обстеження, виникає необхідність у контролі якості ваг.

IV. Порядок розрахунку складових системи статистичних ваг

1. Розрахунок базової ваги домогосподарств

Базова вага i -го домогосподарства при поширенні результатів ОРС на генеральну сукупність розраховується за формулою:

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

2. Урахування відмов домогосподарств від обстеження

1. Для врахування відмов домогосподарств від обстеження будується спеціальна система комірок, яка включає систему 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 й забезпечення певної кількості домогосподарств у кожній комірці.

2. При побудові системи комірок в ОРС використовуються такі класифікаційні ознаки, як регіон і тип місцевості. Крім того, ураховується ступінь територіальної близькості груп домогосподарств, що включені до певної комірки.

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

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

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

6. Коефіцієнт урахування відмов домогосподарств від обстеження розраховується по кожній l -й комірці за формулою:

$$k_l = k_{1i} = \frac{\sum_{i=1}^n w_{Bi} \times \lambda_{li}}{\sum_{i=1}^n w_{Bi} \times \lambda_{li} \times \eta_{li}}, \quad (3)$$

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

M_l – сукупність домогосподарств l -ї комірки;

$\eta_{li} = \begin{cases} 1, \text{ якщо домогосподарство взяло участь в обстеженні;} \\ 0, \text{ якщо домогосподарство відмовилось від обстеження.} \end{cases}$

7. Величини коефіцієнтів урахування відмов домогосподарств від обстеження k_{1i} однакові в межах кожної окремої комірки.

8. Статистична вага i -го домогосподарства з урахуванням відмов домогосподарств від обстеження w_{1i} розраховується за формулою:

$$w_{1i} = w_{Bi} \times k_{1i}. \quad (4)$$

3. Урахування відмов осіб від обстеження

1. Для врахування відмов окремих осіб від обстеження за сукупністю обстежених домогосподарств будується система комірок. Код комірки визначається аналогічно коду комірки для врахування відмов домогосподарств від обстеження.

2. Загальна кількість осіб у комірці n_p визначається за результатами обстеження домогосподарств за поточний місяць. Коефіцієнт урахування відмов осіб від обстеження k_{2r} розраховується по кожній p -й комірці за формулою:

$$k_p = k_{2r} = \frac{\sum_{i=1}^n w_{1i} \times \lambda_{pr}}{\sum_{i=1}^n w_{1i} \times \lambda_{pr} \times \eta_{pr}}, \quad (5)$$

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

M_p – сукупність осіб – членів обстежених домогосподарств p -ї комірки;

$\eta_{pr} = \begin{cases} 1, \text{ якщо особа взяла участь в обстеженні;} \\ 0, \text{ якщо особа відмовилась від участі в обстеженні.} \end{cases}$

3. Величини коефіцієнтів урахування відмов осіб від обстеження k_{2r} однакові в межах кожної окремої комірки.

4. Статистична вага r -ї особи i -го домогосподарства з урахуванням відмов

від інтерв'ю $w_{2i,r}$ розраховується за формулою:

$$w_{2i,r} = w_{1i} \times k_{2r} . \quad (6)$$

4. Урахування статевовікової структури населення

1. Для врахування статевовікової структури населення усуваються постстратифікаційні ефекти. З цією метою будується система комірок за такими класифікаційними ознаками:

регіон – дискретні значення за кількістю регіонів;

тип місцевості – 2 дискретні значення (міська та сільська місцевості);

стать респондента – 2 дискретні значення (жіноча, чоловіча);

вікова група, до якої входить респондент, – 14 дискретних значень (15–19 років, 20–24 роки, 25–29 років, 30–34 роки, 35–39 років, 40–44 роки, 45–49 років, 50–54 роки, 55–59 років, 60–64 роки, 65–70 років, 71–74 років, 75–89 років, 90 років і старше).

2. Код комірки c_q формується так:

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

Код комірки використовується у процедурі коригування системи статистичних ваг, що реалізується в масиві мікроданих ОРС.

3. Загальна кількість осіб у комірці n_q визначається за результатами проведення обстеження за поточний місяць.

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

5. Коефіцієнти для усунення постстратифікаційних ефектів k_{3r} розраховуються по кожній q -й комірці за формулою:

$$k_q = k_{3r} = \frac{N_q}{\sum_{i=1}^n w_{2i,r} \times \lambda_{qr}} , \quad (7)$$

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

M_q – сукупність осіб q -ї комірки.

6. Величини коефіцієнтів для усунення постстратифікаційних ефектів k_{3r}

однакові в межах кожної окремої комірки. Статистичні ваги осіб після усунення постстратифікаційних ефектів w_{3i} розраховуються за формулою:

$$w_{3i,r} = w_{2i,r} \times k_{3r}. \quad (8)$$

Кінцеві статистичні ваги осіб $w_{i,r}$ дорівнюють $w_{3i,r}$.

7. На основі щомісячних статистичних ваг розраховується середня статистична вага за квартал, півріччя, 9 місяців і рік шляхом підсумовування щомісячних даних щодо вагових коефіцієнтів та ділення одержаної суми на відповідну кількість місяців.

8. У додатку 2 до цієї Методики для ілюстрації наведено значення ваг і коефіцієнтів коригування ваг для домогосподарств Любешівського району Волинської області за результатами основного інтерв'ю обстеження робочої сили 2021 року.

V. Порядок визначення якості системи статистичних ваг

1. Якість системи статистичних ваг ОРС визначається після кожного етапу їх коригування, а саме після визначення загальної ймовірності відбору домогосподарств і базової ваги, коригування ваг для врахування ступеня участі відібраних одиниць в обстеженні та постстратифікації статистичних ваг окремо за результатами основного та всіх щомісячних інтерв'ю.

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

3. Під час контролю якості системи статистичних ваг розраховуються такі характеристики:

1) середня вага по всіх домогосподарствах або особах (у залежності від етапу розрахунку системи ваг) як окремий множник – коефіцієнт інфляції F .

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

$$F_{(j)} = \frac{\sum_{i=1}^n w_{(j)i}}{n} \quad (9)$$

2) максимальне значення $w_{(j)\max}$ та мінімальне значення $w_{(j)\min}$ ваги домогосподарств (осіб). При аналізі отриманих результатів мати на увазі, що є неприпустимими мінімальні значення $w_{\min}^{(j)} \leq 0$. За таких умов необхідно перевірити процедури розрахунку системи статистичних ваг і в разі необхідності забезпечити додатні значення величин ваг із застосуванням спеціальних процедур [7];

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

$$R_{(j)1} = \frac{F_{(j)}}{w_{(j)\min}} \quad R_{(j)2} = \frac{w_{(j)\max}}{F_{(j)}} \quad R_{(j)3} = \frac{w_{(j)\max}}{w_{(j)\min}} \quad (10)$$

Ці відношення характеризують ступінь відхилення крайніх значень ваг від середнього значення та між собою у відносних величинах;

4) розмах варіації ваг, як характеристика їх варіативності, розраховується за формулою:

$$\Delta R_{(j)} = w_{(j)\max} - w_{(j)\min} \quad (11)$$

5) відносне зростання дисперсії вибірових оцінок, джерелом яких є варіативність ваг одиниць вибірки, k_L . Величина k_L визначається за формулою:

$$k_L = \frac{n \sum_{i=1}^n (w_{(j)i})^2}{(\sum_{i=1}^n w_{(j)i})^2} \quad (12)$$

Величина k_L характеризує ступінь зростання величини дисперсії вибірових оцінок показників ОРС 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:

$$CV(w_{(j)}) = \sqrt{\frac{1}{n \cdot F_{(j)}^2} \times \sum_{i=1}^n (w_{(j)i} - F_{(j)})^2}, \quad (13)$$

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

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

$$R_{F(j)} = \frac{F_{(j)}}{F_{(j-1)}} \quad (14)$$

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

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

$$R_{(j)4} = \frac{cv(w_{(j)})}{cv(w_{(j-1)})} \quad (15)$$

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

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

$$r_{(j,j-1)} = \frac{\sum_{i=1}^n (w_{(j-1)i} - \bar{w}_{(j-1)}) \times (w_{(j)i} - \bar{w}_{(j)})}{\sqrt{\sum_{i=1}^n (w_{(j-1)i} - \bar{w}_{(j-1)})^2 \times \sum_{i=1}^n (w_{(j)i} - \bar{w}_{(j)})^2}} \quad (16)$$

Коефіцієнт кореляції характеризує щільність лінійного статистичного зв'язку систем статистичних ваг, що відповідають суміжним етапам розрахунку ваг. Що ближче є модуль величини цього коефіцієнта до одиниці, то менше розподіл поточних ваг відхиляється від розподілу попереднього етапу й у цілому від початкових ваг дизайну;

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

$$N_{(j)} = \sum_{i=1}^n w_{(j)i} \quad (17)$$

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

$$D_{(j)} = N_{3I} - N_{(j)} \quad (18)$$

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

11) оцінки загальної кількості населення України та регіонів за розрізами (тип місцевості, стать, вікові групи) розраховуються по масиву первинних даних для осіб, що проживають у домогосподарствах, по яких проведено інтерв'ю з використанням ваг $w_{(j)i}$. Результати порівнюються з даними демографічної статистики.

4. Приклад окремих результатів оцінювання якості статистичних ваг (з урахуванням невідповідей домогосподарств і постстратифікації) наведено в додатку 3 до цієї Методики. Наведені дані ілюструють зміну окремих характеристик якості системи статистичних ваг обстеження робочої сили впродовж 2019–2021 років.

Директор департаменту обстежень
домогосподарств апарату Держстату

Алла СОЛОП

Додаток 1
до Методики
(пункт 4 розділу III)

**Схема визначення системи статистичних ваг для оцінки показників за
результатами обстеження робочої сили**



Додаток 2
до Методики
(пункт 8 підрозділу 4 розділу IV)

Ваги домогосподарств і коефіцієнти коригування статистичних ваг для домогосподарств Любешівського району Волинської області за результатами обстеження робочої сили у 2021 році

Код домогосподарства	w_{Bi}	k_{1i}	w_{1i}	k_{2i}	w_{2i}	k_{3i}	w_{3i}
31255300345	341,63	1,00	341,63	1,01	345,73	1,02	352,64
31255300553	341,63	1,00	341,63	1,01	345,73	1,06	366,47
31256100147	341,63	1,00	341,63	1,04	355,64	1,32	469,44
31256100343	341,63	1,00	341,63	1,04	355,64	1,05	373,42
31256200312	341,63	1,00	341,63	1,02	349,83	1,35	472,27
31256200117	341,63	1,00	341,63	1,02	349,83	1,47	514,25

Додаток 3
до Методики
(пункт 4 розділу V)

**Окремі характеристики якості системи статистичних ваг
обстеження робочої сили у 2019–2021 роках**

Характеристики	2019	2020	2021
З урахуванням невідповідей домогосподарств			
Середня вага, F	1156	1149	1147
Мінімальна вага, $w_{\min}$	16	16	16
Максимальна вага, $w_{\max}$	3910	4832	4998
Коефіцієнт варіації ваг, CV_w , %	50,0	51,1	50,6
$R_1 = \frac{F}{w_{\min}}$	72,25	71,81	71,69
$R_2 = \frac{w_{\max}}{F}$	3,38	4,20	4,36
$R_3 = \frac{w_{\max}}{w_{\min}}$	244,38	302,00	312,38
Розмах варіації ваг, ΔR	3894	4816	4982
З урахуванням постстратифікаційних ефектів			
Середня вага, F	2241	2356	2410
Мінімальна вага, $w_{\min}$	94	161	149
Максимальна вага, $w_{\max}$	26334	27121	22309
Коефіцієнт варіації ваг, CV_w , %	66,4	71,7	73,0
$R_1 = \frac{F}{w_{\min}}$	23,84	14,63	16,17
$R_2 = \frac{w_{\max}}{F}$	11,75	11,51	9,26
$R_3 = \frac{w_{\max}}{w_{\min}}$	280,15	168,45	149,72
Розмах варіації ваг, ΔR	26240	26960	22160

Список використаних джерел

1. Закон України "Про офіційну статистику" від 16 серпня 2022 року № 2524-IX.
URL: <https://zakon.rada.gov.ua/laws/show/2524-20>
2. Вибіркове спостереження: Термінологічний словник./ О. О. Васечко, О. І. Черняк, Є. М. Жуйкова та ін.. – К.: Державний комітет статистики України, 2004. – 140 с.
3. Глосарій до плану статистичного спостереження, затверджений наказом Держкомстату від 29 грудня 2009 року № 498.
URL: http://www.ukrstat.gov.ua/metod_polog/glos.htm
4. Глосарій термінів національної моделі діяльності органів державної статистики, затверджений наказом Держстату від 28 грудня 2022 року № 392.
URL: https://ukrstat.gov.ua/norm_doc/2022/392/glosi_GAMSO.pdf
5. Методологічні положення щодо формування вибірових сукупностей для проведення органами державної статистики України вибірових обстежень населення (домогосподарств), затверджені наказом Держстату від 26 лютого 2018 року № 32.
URL: https://www.ukrstat.gov.ua/metod_polog/metod_doc/2018/32/mp_vond.pdf
6. Методика формування вибірових сукупностей для проведення у 2019–2024 роках вибірових обстежень населення (домогосподарств): умов життя домогосподарств, економічної активності населення та сільськогосподарської діяльності населення у сільській місцевості, затверджена наказом Держстату від 01 березня 2018 року № 39.
URL: https://www.ukrstat.gov.ua/norm_doc/2022/402/402.pdf
7. Саріогло В. Г. Проблеми статистичного зважування вибірових даних : [монографія] / Саріогло В. Г. – К.: ІВЦ Держкомстату України, 2005. – 264 с.
8. Статистичний словник / [О. Г. Осауленко, О. О. Васечко, М. В. Пугачова та ін.]; за ред. д-ра держ. упр., проф., член-кор. НАН України О. Г. Осауленка; НТК стат. досл. – К.: ДП "Інформ.-аналіт. агентство", 2012. – 498 с.
URL: http://194.44.12.92:8080/jspui/bitstream/123456789/2156/1/Slovar_2012-pravkaA4kadr.pdf
9. Regulation (EU) 2019/1700 of the European Parliament and of the Council of 10 October 2019 establishing a common framework for European statistics relating to persons and households, based on data at individual level collected from samples, amending Regulations (EC) № 808/2004, (EC) № 452/2008 and (EC)

№ 1338/2008 of the European Parliament and of the Council, and repealing Regulation (EC) № 1177/2003 of the European Parliament and of the Council and Council Regulation (EC) № 577/98.

URL: <http://data.europa.eu/eli/reg/2019/1700/oj>

10. Commission Implementing Regulation (EU) 2019/2240 of 16 December 2019 specifying the technical items of the data set, establishing the technical formats for transmission of information and specifying the detailed arrangements and content of the quality reports on the organisation of a sample survey in the labour force domain in accordance with Regulation (EU) 2019/1700 of the European Parliament and of the Council.

URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2240&from=EN>

11. Kalton G. Introduction to Survey Sampling. – Beverly Hills: Sage, 1983.

12. Kish L.: Selected papers. – New York: Wiley, 2003.

13. Kish L. Survey sampling. – New York: Wiley, 1995.