

Державна служба статистики України

ЗАТВЕРДЖЕНО

Наказ Державної служби
статистики України

29.12.2017 № 339

**МЕТОДИКА
РОЗРАХУНКУ СИСТЕМИ СТАТИСТИЧНИХ ВАГ ДЛЯ ПОШИРЕННЯ
РЕЗУЛЬТАТІВ ВИБІРКОВОГО ОБСТЕЖЕННЯ НАСЕЛЕННЯ
(ДОМОГОСПОДАРСТВ) З ПИТАНЬ ЕКОНОМІЧНОЇ АКТИВНОСТІ
(ОБСТЕЖЕННЯ РОБОЧОЇ СИЛИ) НА ГЕНЕРАЛЬНУ СУКУПНІСТЬ**

Державна служба статистики України

Відповідальний за випуск – директор департаменту статистики праці
Сеник І. В.

Методика визначає основні принципи та порядок розрахунку системи статистичних ваг домогосподарств і осіб у вибірковому обстеженні населення (домогосподарств) з питань економічної активності.

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

Методику призначено для використання працівниками органів державної статистики, які займаються організацією, проведенням і аналізом результатів вибіркових обстежень населення (домогосподарств). Вона є також корисною для зацікавлених користувачів статистичної інформації, фахівців наукових організацій і навчальних закладів, що проводять дослідження в галузі методології вибіркових обстежень.

Методику підготували працівники департаменту статистики праці
Держстату:

Сеник І. В. – директор департаменту статистики праці;

Солоп А. В. – заступник директора департаменту - начальник відділу комплексних робіт;

Неверовська Л. А. – начальник відділу методології статистики ринку праці;

Самойленко О. С. – головний економіст - спеціаліст методології статистики ринку праці

та фахівець Інституту демографії та соціальних досліджень ім. М. В.Птухи НАН України Саріогло В. Г. – 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 28.12.2017 № 26).

Державна служба статистики України

- адреса: вул. Шота Руставелі 3, Київ, 01601
- телефон: (044) 287-24-22
- факс: (044) 235-37-39
- електронна пошта: office@ukrstat.gov.ua
- веб-сайт: www.ukrstat.gov.ua

Зміст

	Стор.
Передмова	4
I. Загальні положення	5
II. Визначення термінів і основних понять.....	6
III. Принципи оцінювання показників	7
IV. Порядок розрахунку складових системи ваг	12
1. Розрахунок базової ваги домогосподарств.....	12
2. Урахування відмов домогосподарств від обстеження	12
3. Урахування відмов осіб від обстеження.....	13
4. Урахування статевовікової структури	14
5. Визначення статистичних ваг за квартал, півріччя, 9 місяців та рік.....	15
V. Порядок визначення якості системи ваг	15
VI. Інформаційна база розрахунку системи статистичних ваг.....	18
Додатки:	
Додаток 1. Порядок визначення системи ваг для оцінки показників за результатами обстеження робочої сили.....	20
Додаток 2. Ваги домогосподарств і коефіцієнти коригування ваг для домогосподарств Любешівського району Волинської області за результатами обстеження робочої сили у 2016 році	21
Додаток 3. Окремі характеристики якості системи ваг обстеження робочої сили у 2012–2016 роках.....	22
Список використаних джерел	23

Передмова

Методика розрахунку системи статистичних ваг для поширення результатів вибіркового обстеження населення (домогосподарств) з питань економічної активності (обстеження робочої сили) на генеральну сукупність (далі – Методика) розроблена з метою визначення основних принципів поширення результатів вищезазначеного обстеження на генеральну сукупність, порядку розрахунку складових системи ваг та визначення їх якості.

Ця Методика є складовою статистичних метаданих державного статистичного спостереження "Економічна активність населення". Інші складові, зокрема, питання методології та організації спостереження, формування вибіркової сукупності, розрахунку характеристик надійності оцінювання показників, класифікації та аналізу показників за результатами спостереження висвітлено у відповідних методологічних положеннях та методиках.

При підготовці Методики враховано сучасні методологічні засади, які використовуються у міжнародній практиці, а також зміни, які відбулися при формуванні вибірки для вибіркового обстеження населення (домогосподарств) з питань економічної активності.

Методика запроваджується в дію на заміну Методики розрахунку системи ваг для поширення результатів вибіркового обстеження населення (домогосподарств) з питань економічної активності на генеральну сукупність, затвердженої наказом Державного комітету статистики України від 14 листопада 2006 р. № 520.

Документ розроблено на виконання заходів Стратегії розвитку державної статистики на період до 2017 року, затвердженої розпорядженням Кабінету Міністрів України від 20 березня 2013 р. № 415-р.

I. Загальні положення

1. Поширення результатів вибіркового обстеження населення (домогосподарств) з питань економічної активності (далі – обстеження робочої сили) на генеральну сукупність здійснюється шляхом реалізації процедури їх статистичного зважування.

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

3. Можливість дотримання принципу імовірності в обстеженні робочої сили 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. Відповідно для забезпечення репрезентативності результатів зазначеного обстеження 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).

4. З 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 із інформацією зовнішніх джерел (зокрема із застосуванням процедур калібрації). Вона дозволяє об'єднувати результати різних вибірових обстежень на нижчих рівнях агрегації даних та на мікрорівні.

5. Система ваг є невід'ємною складовою результатів обстеження робочої сили. Її розрахунок здійснюється з метою визначення результуючої ваги кожної одиниці спостереження, що була обстежена, та оцінювання показників по генеральній сукупності за результатами обстеження.

II. Визначення термінів і основних понять

1. У цій Методиці терміни вживаються у значеннях, визначених Глосарієм до плану статистичного спостереження [2], а також у такому значенні:

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

2) вибіркоче обстеження населення (домогосподарств) – науково обґрунтований вид несучільного спостереження, при якому обстежуються не всі одиниці сукупності, а лише певним чином відібрана їх частина, а одиницею спостереження є особа чи домогосподарство;

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

4) генеральна сукупність – множина одиниць (неінституційні домогосподарства України), характеристики яких необхідно дослідити і щодо яких буде зроблене узагальнення за результатами вибіркового обстеження;

5) імовірність – число, яке характеризує ступінь можливості певного елемента генеральної сукупності потрапити до вибірки;

6) інституційне населення – це військовослужбовці строкової служби; особи, які знаходяться в місцях позбавлення волі; особи, які постійно проживають у будинках-інтернатах, будинках для осіб похилого віку тощо;

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

8) 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;

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

10) план (дизайн) вибірки – організаційно-логічна модель структури вибіркової сукупності та принципи її формування. Якість плану вибірки

підтверджує репрезентативність отриманої за визначеними у плані правилами вибіркової сукупності;

11) постстратифікація – трансформація (коригування) системи ваг одиниць спостереження шляхом забезпечення відповідності оцінок показників за результатами вибіркового обстеження зовнішній інформації по певних стратах генеральної сукупності;

12) репрезентативність – властивість вибіркової сукупності повно та адекватно відображати основні характеристики генеральної сукупності;

13) статистична вага – величина, обернена до ймовірності відбору одиниці вибірки;

14) стратифікація – поділ генеральної сукупності (або її частини) на групи, однорідні за певною ознакою;

15) 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).

III. Принципи оцінювання показників

1. Обстеження робочої сили є вибірконим 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, у тому числі й обстеження робочої сили.

2. 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 імовірностей відбору одиниць вибірки, фактичних рівнів відмов домогосподарств та окремих осіб від участі в обстеженні, гармонізації результатів обстеження з даними демографічної статистики щодо кількості та статеві-вікової структури населення. Таким чином, система статистичних ваг відображає дизайн вибірки, а також ураховує основні етапи реалізації обстеження.

3. Генеральну сукупність обстеження робочої сили складають усі особи у віці 15–70 років (з 2019 року – 15 років і старше), які постійно проживають у неінституційних домогосподарствах.

4. Інформаційною базою для формування основи вибірки обстеження робочої сили у 2014–2018 роках є підготовчі матеріали Всеукраїнського перепису населення "Складання квартальних списків житлових будинків та житлових приміщень в інших будівлях у міських поселеннях та великих сільських населених пунктах та списків сільських населених пунктах".

Для формування вибірки на 2019–2023 роки будуть використовуватися дані Державного реєстру виборців щодо розподілу окремих територій на виборчі дільниці та інформація щодо окремих характеристик цих дільниць.

5. Порядок формування вибіркової сукупності для проведення обстеження робочої сили один раз на п'ять років викладається у методиці формування вибірових сукупностей для проведення вибірових обстежень населення (домогосподарств): умов життя домогосподарств, економічної активності населення та сільськогосподарської діяльності населення у сільській місцевості (далі – методика щодо формування вибірових сукупностей). Методика щодо формування вибірових сукупностей на 2014–2018 роки затверджена наказом Держстату від 08 лютого 2013 р. № 37.

6. Для поширення результатів обстеження робочої сили на генеральну сукупність розраховується система статистичних ваг обстежених осіб. Статистичні ваги призначені для врахування загальних імовірностей уключення домогосподарств до вибірки, урахування фактичних рівнів відмов домогосподарств та окремих осіб від участі в обстеженні, гармонізації результатів обстеження з даними демографічної статистики щодо чисельності та статевовікової структури населення тощо.

7. Розрахунок системи статистичних ваг складається з таких основних етапів (додаток 1):

- 1) розрахунок базових ваг домогосподарств;
- 2) коригування базових ваг для врахування відмінностей між теоретичною та фактичною кількостями відібраних домогосподарств;
- 3) коригування ваг для врахування ступеня участі домогосподарств в обстеженні;
- 4) коригування ваг для врахування ступеня участі осіб в обстеженні;
- 5) коригування ваг для узгодження результатів обстеження з наявною інформацією із зовнішніх джерел та визначення кінцевої ваги кожного

обстеженого домогосподарства та особи;

б) контроль якості системи ваг.

8. У загальному випадку кінцева вага r -ї особи розраховується як добуток базової ваги домогосподарства та відповідних вагових коефіцієнтів (коефіцієнтів переважування) за формулою:

$$w_r = w_{Br} \cdot k_{1r} \cdot \dots \cdot k_{Er}, \quad (1)$$

де w_r – кінцева вага особи;

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

$k_{1r} \cdot \dots \cdot k_{Er}$ – вагові коефіцієнти;

E – кількість етапів коригування статистичних ваг.

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

9. Базова вага є результатом певної послідовності розрахунків. Реалізований в обстеженні дизайн вибірки зумовлює використання для побудови базових ваг по кожній страті таких складових:

1) імовірність відбору ПТОВ – P_1 ;

2) імовірність відбору ВТОВ по міських поселеннях або домогосподарств у сільській місцевості – P_2 ;

3) імовірність відбору домогосподарств у межах ВТОВ по міських поселеннях – P_3 .

Імовірності P_1, P_2, P_3 визначаються за результатами формування вибірки обстеження робочої сили.

10. В ймовірнісній вибірці обстеження робочої сили кожне i -те домогосподарство генеральної сукупності обсягу N , $i = 1, 2, \dots, N$ повинне мати певну позитивну ймовірність 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.

11. Базові ваги розраховуються при формуванні вибірки на весь термін її експлуатації. Територіальна вибірка для обстеження робочої сили розрахована на п'ять років.

12. Якість результатів обстеження робочої сили може суттєво погіршуватися внаслідок впливу такого явища, як відмова 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 з лише з таких осіб. Особливо це стосується територій зі значними рівнями відмов.

13. Одним з найбільш розповсюджених і теоретично обґрунтованих методів 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.

14. Коригування ваг для компенсації відмов уключає такі кроки:

1) розділення вибірки на систему комірок – класів зважування, що охоплюють усю вибірову сукупність і не перетинаються. Комірки відповідають окремим ВТОВ, ПТОВ або формуються об'єднанням декількох сусідніх ВТОВ або ПТОВ так, щоб загальна кількість обстежених домогосподарств у комірці була не меншою за встановлене граничне значення (у поточній вибірці гранична мінімальна кількість домогосподарств у комірці становить 30);

2) розрахунок єдиних коефіцієнтів коригування ваг для всіх одиниць комірки, як величини оберненої до частки одержаних відповідей від домогосподарств та осіб;

3) контроль коефіцієнтів коригування ваг з метою запобігання дуже значним коригуванням ваг;

4) коригування системи ваг домогосподарств та осіб на основі визначених коефіцієнтів.

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

16. Доцільність узгодження результатів обстеження з інформацією із зовнішніх джерел зумовлена такими обставинами:

1) оцінки показників за результатами вибірового обстеження завжди характеризуються певною похибкою, джерелами якої є несучільність

обстеження (похибка вибірки) і наявність так званих невібіркових похибок;

2) 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;

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

17. З відомих на сьогодні узагальнених методів коригування системи статистичних ваг найбільш теоретично розробленою й ефективною є так звана процедура калібрації статистичних ваг. Термін "калібрація" означає трансформацію (коригування) системи ваг одиниць вибірки шляхом мінімізації функції відстані (відмінностей) між калібраційними вагами та вагами, що коригуються, за умови, що калібраційні ваги дають точні оцінки зовнішніх даних, відомих для генеральної сукупності або її частин.

18. Метод постстратифікації системи статистичних ваг є спрощеним варіантом калібрації і 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 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.

19. Для постстратифікації статистичних ваг в обстеженні робочої сили використовуються, насамперед, дані статистики населення щодо кількості, розміщення та статевовікової структури населення (далі – дані демографічної статистики).

20. 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. Враховуючи 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 ваг

1. Розрахунок базової ваги домогосподарств

1. Базова статистична вага одиниці вибірки при поширенні результатів обстеження робочої сили на генеральну сукупність розраховується за формулою:

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

2. При формуванні вибіркової сукупності домогосподарств для обстеження робочої сили у 2016 році частка відбору дорівнювала $f = \frac{n}{N} = \frac{16014}{13801945}$ (де n – кількість домогосподарств, відібраних із ПТОВ; N – загальна кількість домогосподарств у ПТОВ). Ця величина розрахована як відношення обсягу вибіркової сукупності обстеження робочої сили за місяць, визначена з урахуванням оптимального рівня надійності показників за результатами обстеження, вартості обстеження та очікуваної участі домогосподарств та осіб в обстеженні до оцінки загальної кількості неінституційних домогосподарств в Україні. Таким чином, середньомісячна базова вага домогосподарств в обстеженні робочої сили складає $w_{Bi} = \frac{1}{f} \approx 862$.

2. Урахування відмов домогосподарств від обстеження

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

2. При побудові системи комірок в обстеженні робочої сили використовуються такі класифікаційні змінні, як регіон та тип місцевості. Крім того, враховується 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 міських поселеннях і ПТОВ у сільській місцевості.

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

5. Загальна кількість домогосподарств у комірці n_l 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:

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

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;

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

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

7. Ваги домогосподарств із урахуванням відмов домогосподарств від обстеження w_{1i} розраховуються за формулою:

$$w_{1i} = w_{Bi} \cdot k_{1r} \quad (4)$$

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 проведення інтерв'ю за поточний місяць. Коефіцієнт урахування відмов осіб від обстеження k_{2r} розраховується по кожній p -й комірці за формулою:

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

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

M_p – сукупність осіб p -ї комірки;

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

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

3. Ваги осіб з урахуванням відмов від інтерв'ю w_{2i} розраховуються за формулою:

$$w_{2i} = w_{1i} \cdot k_{2r} \quad (6)$$

4. Урахування статевовікової структури населення

1. Для урахування статевовікової структури населення здійснюється усунення постратифікаційних ефектів. З цією метою будується система комірок за такими класифікаційними змінними:

- 1) регіон – дискретні значення за кількістю регіонів;
- 2) тип місцевості – 2 дискретні значення (міські поселення, сільська місцевість);
- 3) стать респондента – 2 дискретні значення (жіноча, чоловіча);
- 4) вікова група, до якої входить респондент – 11 дискретних значень (15–19 років, 20–24 роки, 25–29 років, 30–34 роки, 35–39 років, 40–44 роки, 45–49 років, 50–54 роки, 55–59 років, 60–64 роки, 65–70 років), а з 2019 року – 13 дискретних значень – додатково 70–74 роки, 75 і старше років.

2. Для ідентифікації комірки розраховується код комірки c_q за формулою:

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

Код комірки використовується у процедурі коригування системи ваг, що реалізується у масиві мікроданих обстеження робочої сили.

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 на 1 січня поточного року.

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

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

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

M_q – сукупність осіб q -ї комірки.

6. Величини коефіцієнтів для усунення постстратифікаційних ефектів k_{3r} однакові в межах кожної окремої комірки. Ваги осіб після усунення постстратифікаційних ефектів w_{3i} розраховуються за формулою:

$$w_{3i} = w_{2i} \cdot k_{3r} \quad (8)$$

Кінцеві ваги осіб w_r дорівнюють w_{3i} .

7. У додатку 2 для ілюстрації наведено значення ваг і коефіцієнтів коригування ваг для домогосподарств Любешівського району Волинської області за результатами основного інтерв'ю обстеження робочої сили 2016 року.

5. Визначення статистичних ваг за квартал, півріччя, 9 місяців та рік

1. На основі щомісячних ваг розраховується середня вага за квартал, півріччя, 9 місяців та рік шляхом підсумовування місячних даних щодо вагових коефіцієнтів та ділення одержаної суми за відповідну кількість місяців.

V. Порядок визначення якості системи ваг

1. Якість системи ваг обстеження робочої сили 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 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 ваг як правило на практиці 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 неприпустимим значення $w_{\min}^{(j)} \leq 0$. 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) співвідношення між коефіцієнтом інфляції та мінімальним і максимальним значеннями ваг – $R_{(j)1}$, $R_{(j)2}$ та $R_{(j)3}$ розраховуються за формулами:

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

Ці відношення характеризують ступінь відхилення крайніх значень ваг від середнього значення та між собою у відносних величинах. Значення $R_{(j)1}$ та $R_{(j)2}$ характеризують ступінь розподілу ваг;

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 [8]:

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

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

б) коефіцієнт варіації ваг $CV(w_{(j)})$ розраховується за формулою:

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

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 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:

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

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 розраховується за формулою:

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

Коефіцієнт кореляції 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) 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 використовувалась 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 чисельності населення за розрізами регіонів (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 ваг (з урахуванням невідповідей домогосподарств і постстратифікації) наведено в додатку 3. Наведені дані ілюструють зміну окремих характеристик якості системи ваг обстеження робочої сили протягом 2012–2016 років.

VI. Інформаційна база розрахунку системи статистичних ваг

1. Інформаційна база для реалізації процедур розрахунку системи ваг за результатами щомісячного обстеження робочої сили включає:

характеристики вибіркової сукупності домогосподарств;
результати обстеження робочої сили;
інформацію із зовнішніх джерел.

2. Характеристики вибіркової сукупності передбачають дані щодо дизайну та основних параметрів вибірки, а саме:

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

3. Характеристики вибіркової сукупності залишаються незмінними протягом усього терміну дії територіальної вибірки для обстеження робочої сили, 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. За результатами обстеження робочої сили отримуються:

масиви обстежених домогосподарств і осіб за результатами поточного обстеження робочої сили, а саме: код домогосподарства, код особи, стать особи, вік особи, статус економічної активності особи, загальнодержавний номер територіальної одиниці;

номер комірки для врахування відмов домогосподарств від обстеження, l ;

код комірки, c_l ;

загальна початкова кількість домогосподарств у комірці, n_l ;

номер комірки для врахування відмов осіб від обстеження, p ;

код комірки, c_p ;

загальна початкова кількість осіб у комірці, n_p ;

номер комірки для усунення постстратифікаційних ефектів, q ;

код комірки, c_q ;

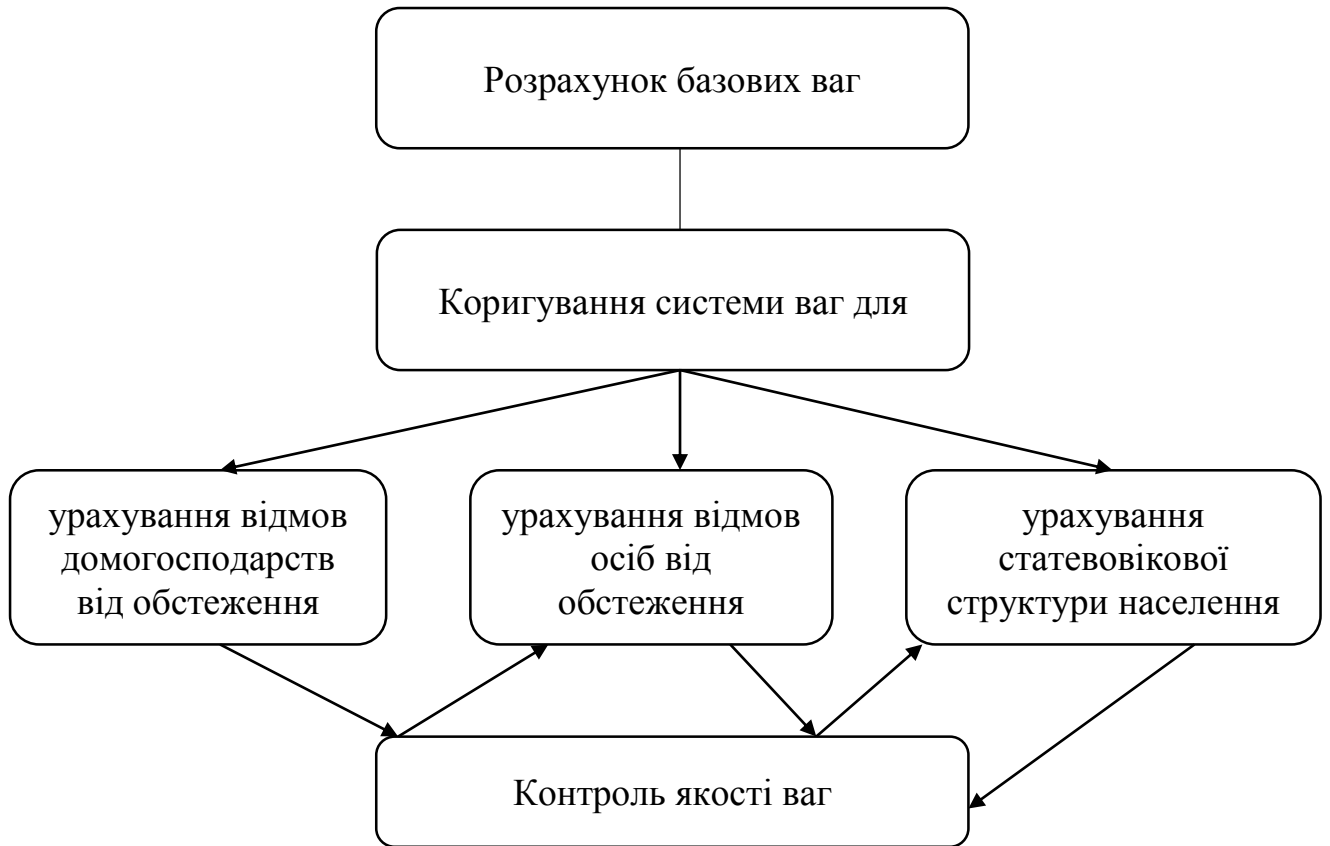
загальна початкова кількість осіб у комірці, n_q .

5. Дані демографічної статистики щодо статевовікової структури населення за п'ятирічними віковими групами на початок поточного року.

Директор департаменту
статистики праці Держстату

І. В. Сенік

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



Директор департаменту
статистики праці Держстату

І. В. Сенік

Додаток 2
до Методики
(пункту 7 розділу IV)

**Ваги домогосподарств і коефіцієнти коригування ваг для домогосподарств
Любешівського району Волинської області за результатами обстеження
робочої сили у 2016 році**

Код домогосподарства	w_{Vi}	k_{1i}	w_{1i}	k_{2i}	w_{2i}	k_{3i}	w_{3i}
3176270183	617,61	1,038	641,153	1,008	646,50	1,02	658,99
3176270446	617,61	1,038	641,153	1,008	646,50	1,56	1009,32
3176250261	617,61	1,036	640,119	1,017	650,88	1,64	1070,49
3176260056	617,61	1,036	640,119	1,017	650,88	1,27	825,36
3176280142	617,61	1,036	639,795	1,008	644,95	1,56	1005,64
3176390018	617,61	1,036	639,795	1,008	644,95	1,20	775,67

Директор департаменту
статистики праці Держстату

І. В. Сенік

Додаток 3
до Методики
(пункт 4 розділу V)

**Окремі характеристики якості системи ваг
обстеження робочої сили у 2012-2016 роках
(з урахуванням невідповідей домогосподарств)**

Характеристики	2012	2013	2014	2015	2016
Середня вага, F	892	892	1275	1171	1174
Мінімальна вага, w_{\min}	88	88	102	67	67
Максимальна вага, w_{\max}	4829	5333	5753	6541	7926
Коефіцієнт варіації ваг, CV_w , %	51,5	51,5	52,2	59,8	61,3
$R_1 = \frac{F}{w_{\min}}$	10,14	10,14	12,50	11,48	17,52
$R_2 = \frac{w_{\max}}{F}$	5,41	5,98	4,51	5,59	6,75
$R_3 = \frac{w_{\max}}{w_{\min}}$	54,87	60,60	56,40	97,63	118,30
Розмах варіації ваг, ΔR	4741	5245	5651	6475	7859

**Окремі характеристики якості системи ваг
обстеження робочої сили у 2012-2016 роках
(з урахуванням постстратифікаційних ефектів)**

Характеристики	2012	2013	2014	2015	2016
Середня вага, F	1227	1233	1247	1148	1176
Мінімальна вага, w_{\min}	98	97	96	90	96
Максимальна вага, w_{\max}	8261	21909	57981	16209	15046
Коефіцієнт варіації ваг, CV_w , %	64,1	65,4	84,4	76,3	79,3
$R_1 = \frac{F}{w_{\min}}$	12,52	12,71	12,99	12,75	12,25
$R_2 = \frac{w_{\max}}{F}$	6,73	17,77	46,50	14,12	12,79
$R_3 = \frac{w_{\max}}{w_{\min}}$	84,30	225,87	603,97	180,10	156,73
Розмах варіації ваг, ΔR	8163	21812	57885	16119	14950

Директор департаменту
статистики праці Держстату

І. В. Сенік

Список використаних джерел

1. Вибіркове спостереження: термінол. словн. / Укл.: О. О. Васечко, О. І. Черняк, Є. М. Жуйкова, О. В. Гончар та ін. – К. : Держкомстат України, 2004. – 140 с.
2. Глосарій до плану статистичного спостереження / Державна служба статистики України. – 2010. – 13 с.
3. Економічна активність населення України у 2016 році: Стат. збірник. – К. : Державна служба статистики, 2017.
4. Методологічні основи формування вибірових сукупностей для проведення органами державної статистики України базових державних вибірових обстеження населення (домогосподарств), затверджені наказом Держкомстату від 02.08.2005 № 223.
5. Методика формування вибірових сукупностей для проведення у 2014–2018 роках вибірових обстежень населення (домогосподарств): умов життя домогосподарств, економічної активності населення та сільськогосподарської діяльності населення у сільській місцевості, затверджена наказом Держстату від 08.02.2013 № 37.
6. Методика розрахунку системи статистичних ваг для поширення результатів державного вибірового обстеження умов життя домогосподарств на генеральну сукупність, затверджена наказом Держкомстату від 14.11.2006 № 521.
7. Методика розрахунку системи статистичних ваг для поширення результатів державного вибірового обстеження населення (домогосподарств) з питань економічної активності на генеральну сукупність, затверджена наказом Держкомстату від 14.11.2006 № 520.
8. Наказ Державної служби статистики України від 30.08.2013 № 263 "Про затвердження форм державного статистичного вибірового обстеження населення (домогосподарств) з питань економічної активності".
9. Саріогло В. Г. Проблеми статистичного зважування вибірових даних : [монографія] / Саріогло В. Г. – К. : ІВЦ Держкомстату України, 2005. – 264 с.
10. 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 статистичних досліджень. – К. : ДП "Інформ.-аналіт. агентство", 2012. – 498 с.
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.