

Державна служба статистики України

ЗАТВЕРДЖЕНО

**Наказ Державної служби
статистики**

14 січня 2021 р. № 17

**МЕТОДОЛОГІЧНІ ПОЛОЖЕННЯ З
ФОРМУВАННЯ ВИБІРКОВИХ СУКУПНОСТЕЙ РЕСПОНДЕНТІВ
У РАМКАХ СИСТЕМИ ПРОВЕДЕННЯ ВИБІРКОВИХ ОБСТЕЖЕНЬ
НАСЕЛЕННЯ З УРАХУВАННЯМ ВИБІРКОВОГО ОБСТЕЖЕННЯ
"СТАТИСТИКА ДОХОДІВ І УМОВ ЖИТТЯ В ЄВРОПЕЙСЬКОМУ
СОЮЗІ (EU-SILC)"**

Київ – 2020

Державна служба статистики України

Відповідальний за випуск – директор департаменту обстежень домогосподарств
Осипова І. І.

У Методологічних положеннях з формування вибірових сукупностей респондентів у рамках системи проведення вибірових обстежень населення з урахуванням вибірового обстеження "Статистика доходів і умов життя в Європейському Союзі" (далі – Методологічні положення) наводяться принципи формування ймовірнісних стратифікованих вибірок, у тому числі багатоступеневих територіальних вибірок, для проведення вибірових обстежень населення, зокрема, умов життя домогосподарств, робочої сили, сільськогосподарської діяльності населення в сільській місцевості та обстеження домогосподарств "Статистика доходів і умов життя в Європейському Союзі" (далі – вибірові обстеження населення). У Методологічних положеннях міститься опис особливостей формування та актуалізації основ територіальної вибірки та вибірок респондентів, висвітлюються загальні принципи та схема організації робіт з формування вибірок територій та домогосподарств із урахуванням упровадження обстеження домогосподарств "Статистика доходів і умов життя в Європейському Союзі" (далі – EU-SILC) у національну систему вибірових обстежень населення.

Методологічні положення підготували фахівці департаменту обстежень домогосподарств Державної служби статистики України:

Осипова І. І. – директор департаменту обстежень домогосподарств;

Передерій А. Б. – начальник відділу з організації вибірових обстежень населення;

Кольва Ю. О. – заступник начальника відділу з організації вибірових обстежень населення,

на основі рекомендацій експерта з питань формування вибірових сукупностей для проведення вибірових обстежень населення проєкту Світового банку "Розвиток системи вибірових обстежень населення відповідно до міжнародних норм і стандартів на основі використання сучасних технологій електронного збору даних і способів поширення інформації, а також удосконалення спроможності ДССУ та методологічна підтримка впровадження обстеження домогосподарств "Статистика доходів і умов життя в Європейському Союзі" (EU-SILC)" та за безпосередньою участю Сариоґла В. Г. – д.е.н., завідувача відділу моделювання соціально-економічних процесів і структур Інституту демографії та соціальних досліджень ім. М. В. Птухи НАН України.

Методологічні положення призначені для використання працівниками органів державної статистики, а також можуть бути корисними для фахівців міністерств, інших органів виконавчої влади, наукових організацій та установ, які проводять дослідження з питань методології вибірових обстежень населення, інших зацікавлених користувачів.

Методологічні положення схвалені Комісією з питань удосконалення методології та звітної документації Державної служби статистики України (протокол від 18 грудня 2020 року № КПУМ/27-20).

Держстат

вул. Шота Руставелі 3, Київ, 01601

[http:// www.ukrstat.gov.ua](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.....	6
II Загальні положення.....	9
III Організація робіт із формування вибірки.....	10
IV Визначення вимог до вибірки відповідно до цілей вибіркового обстеження.....	12
V Формування основи вибірки з урахуванням підсумків аналізу якості наявних інформаційних джерел.....	15
VI Розрахунок обсягу вибіркової сукупності.....	16
VII Розробка документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств)".....	22
VIII Розробка плану (дизайну) вибірки	24
IX Реалізація робіт із формування вибірки в рамках системи державних вибірових обстежень населення (домогосподарств).....	28
1. Відбір первинних територіальних одиниць вибірки.....	28
2. Відбір вторинних територіальних одиниць вибірки.....	35
3. Відбір домогосподарств.....	36
X Актуалізація основи вибірки.....	38
Додатки:	
Додаток 1. Макет документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств)".....	41
Додаток 2. Схема ротації домогосподарств в обстеженні робочої сили	42
Додаток 3. Приклад схеми ротації вибірки в обстеженні EU-SILC...	44
Додаток 4. Приклад оцінки обсягу вибірки обстеження EU-SILC в Україні.....	47
Додаток 5. Приклад матеріалів та даних для формування документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірового обстеження EU-SILC"	52
Додаток 6. Приклад розрахунку порога саморепрезентативності.....	58
Додаток 7. Приклад розрахунку мінімального розміру ВТОВ для проведення на окремих мережах респондентів двох вибірових обстежень населення (домогосподарств).....	59
Додаток 8. Приклад упорядкування територіальних одиниць за розміром та за принципом "географічного серпантину", а також реалізації процедури відбору територіальних одиниць.....	61
Список використаних джерел.....	64

Скорочення та умовні позначення

1. У цих Методологічних положеннях використовуються такі умовні позначення:

ПТОВ	–	первинна територіальна одиниця вибірки;
ВТОВ	–	вторинна територіальна одиниця вибірки;
ОУЖД	–	обстеження умов життя домогосподарств;
ОРС	–	обстеження робочої сили;
ОСГД	–	обстеження сільськогосподарської діяльності населення в сільській місцевості;
EU-SILC	–	обстеження домогосподарств "Статистика доходів і умов життя в Європейському Союзі".

2. У цих Методологічних положеннях використовуються символи в таких значеннях:

n	–	обсяг вибірки;
σ^2	–	дисперсія;
q	–	кількість ротаційних груп;
p	–	оцінка очікуваного значення частки;
N	–	загальна кількість населення/домогосподарств (обсяг генеральної сукупності);
m	–	навантаження фахівця з інтерв'ювання;
d	–	кількість одиниць для відбору;
W	–	частка загальної кількості постійного населення регіону;
P	–	імовірність відбору ПТОВ;
CV	–	коефіцієнт варіації;
$\bar{y}$	–	оцінка очікуваного середнього значення показника;
$deff$	–	значення дизайн-ефекту;
K	–	ваговий коефіцієнт;
k_{rr}	–	рівень участі домогосподарств;
rnd	–	випадкове число;
SE	–	стандартна похибка;
$h^{(z)}$	–	накопичене значення ознаки для відбору територіальної одиниці z ;
t	–	кількість років перебування територіальної одиниці у вибірці;
M	–	кількість пронумерованих для відбору домогосподарств у межах територіальної одиниці (розмір територіальної одиниці);
D	–	загальна кількість домогосподарств територіальних одиниць;
c_f	–	кількість одиниць спостереження (осіб, домогосподарств), які взяли участь в обстеженні;
c	–	кількість одиниць спостереження, які були відібрані для участі в обстеженні;
$\hat{Y}$	–	оцінка очікуваного сумарного значення показника;
n_p^{HH}	–	поріг саморепрезентативності (кількість домогосподарств);

- n_p^{pers} – поріг саморепрезентативності (кількість населення (осіб));
- z – розмір ротаційних груп (підгруп) (домогосподарств або осіб);
- d – кількість несаморепрезентативних територіальних одиниць, яку необхідно відібрати;
- b – кількість ВТОВ, які необхідно відібрати;
- f – частка відбору;
- s – середній розмір домогосподарства (кількість осіб).

I. Вступ

1. Метою розробки Методологічних положень є створення методологічної основи розробки плану (дизайну) вибірок для проведення органами державної статистики вибіркових обстежень населення з урахуванням упровадження EU-SILC у національну систему вибіркових обстежень населення, визначення основних етапів організації, проведення та контролю якості робіт із формування вибірок територій та респондентів.

2. Методологічні положення підготовлено з урахуванням рекомендацій міжнародних експертів: Г. Колтона (США) [12, 13], В. Верми [20], В. Саріогла - експерта компонента № CQ-02 "Розвиток системи вибіркових обстежень населення відповідно до міжнародних норм і стандартів на основі використання сучасних технологій електронного збору даних і способів поширення інформації, а також удосконалення спроможності ДССУ та методологічна підтримка впровадження обстеження домогосподарств "Статистика доходів і умов життя в Європейському Союзі" (EU-SILC)" проєкту Світового банку № TF0A7020 "Розробка статистичних методик і розбудова спроможностей в умовах збройного конфлікту в Україні" [23], положень, викладених у документі "Методологія європейських досліджень робочої сили: Приклад проектування та реалізації" Ф. Гагніарді, В. Верма, Д. Чіпапініні [19], а також викладених у Керівництві "Обстеження бюджетів домогосподарств в ЄС. Методологія та рекомендації з гармонізації" (2003) [16], Регламенту (ЄС) № 2019/1700 від 10 жовтня 2019 року "Формування загальної бази європейської статистики стосовно осіб та домогосподарств на основі даних на індивідуальному рівні, зібраних за результатами вибіркових обстежень" [21] та звіті Євростату "Обстеження робочої сили в країнах ЄС, країнах-кандидатах. Основні характеристики національних оглядових показників 2015 року" [15].

3. Методологічні положення враховують досвід проведення Держстатом вибіркових обстежень населення у період 1999–2020 років.

4. Методологічні положення ґрунтуються на нормах Закону України "Про державну статистику" [1].

6. Для цілей Методологічних положень терміни вживаються у значеннях, визначених глосарієм до плану статистичного спостереження [6], викладених у Статистичному словнику [7], а також у таких значеннях:

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

2) вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств) – основні параметри вибірової сукупності та вимоги до її формування відповідно до мети дослідження;

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

4) домогосподарство з вибірки (застосовується в EU-SILC) – домогосподарство з вибірки включається до EU-SILC для збору інформації, якщо його членом є принаймні одна особа з вибірки віком 16 років або старше;

5) звітний період – період часу, до якого відносяться дані;

6) злиття осіб з вибірки (застосовується в EU-SILC) – об'єднання осіб з вибірки з різних домогосподарств із вибірки попереднього циклу з метою утворення нового домогосподарства;

7) колективне домогосподарство (застосовується в EU-SILC) – неінституційне колективне житло, таке як пансіонат, гуртожиток в освітньому закладі чи інше житлове приміщення, в якому спільно проживає більше п'яти осіб, і витрати на утримання домогосподарства в них не є спільними; до таких домогосподарств також відносяться особи, що проживають як квартиронаймачі в домогосподарствах, що мають більше п'яти квартиронаймачів;

8) обсяг вибірки досягнутий – кількість одиниць спостереження у вибіровому обстеженні, по яких зібрані дані, придатні для подальшої обробки й аналізу;

9) обсяг вибірки ефективний – обсяг вибірки, визначений для обстеження, в якому реалізується простий випадковий відбір одиниць спостереження (домогосподарств або осіб) і відповідно дизайн-ефект дорівнює 1,0;

10) обсяг вибірки початковий (фактичний) – кількість одиниць спостереження, яку необхідно відібрати для забезпечення встановленого обсягу ефективної вибірки;

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

12) первісне/відокремлене домогосподарство (застосовується в EU-SILC) – домогосподарство з вибірки циклу x вважається "відокремленим", якщо його особи з вибірки з циклу x проживають на час проведення циклу $x+1$ у більш ніж

одному приватному домогосподарстві в межах території, включеної до сукупності; після відокремлення одне і тільки одне з утворених домогосподарств відповідає визначенню "первісне" домогосподарство, у той час як одне або більше інших домогосподарств називаються "відокремленими" домогосподарствами;

13) період збору даних – період часу, коли збираються дані обстеження;

14) повне домогосподарство (застосовується в EU-SILC) – домогосподарство з вибірки вважається повним, якщо воно протягом панельного обстеження залишається одним домогосподарством без утворення додаткового домогосподарства і без його зникнення, при цьому в ньому можуть відбуватися зміни у складі, порівняно з попереднім циклом, через смерть, вихід членів за межі області дослідження або переїзд членів домогосподарства, приєднання осіб до домогосподарства або народження;

15) поздовжні дані (EU-SILC, OPC) – дані, що відносяться до змін на рівні фізичних осіб протягом тривалого часу і періодично спостерігаються впродовж певного періоду; поздовжні дані можуть надходити або з поперечного обстеження з ротаційною вибіркою, у якій відстежуються відібрані особи, або з панельного обстеження;

16) поперечні (cross-sectional) дані (EU-SILC, OPC) – дані, що відносяться до певного моменту часу або певного часового періоду; поперечні дані можуть бути отримані або з вибіркового обстеження, яке проводиться як поперечне – одноразове, щорічне, щоквартальне і т.ін. (без ротації або з ротацією одиниць спостереження у вибірці), або з панельного вибіркового обстеження за умови забезпечення поперечної репрезентативності;

17) приватне домогосподарство – визначається як особа, що проживає одна, або група людей, що проживають разом в одному приватному житлі та мають спільні витрати, у тому числі витрати на спільне забезпечення предметами першої необхідності; члени такого домогосподарства можуть об'єднувати певною мірою свої доходи; (фактично це визначення відповідає визначенню домогосподарства у статті 1 Закону України "Про Всеукраїнський перепис населення");

18) ризик бідності (за критерієм ЄС) – рівень бідності, визначений за таким відносним критерієм: середньодушові еквівалентні доходи особи є нижчими 60% медіанного рівня середньодушових еквівалентних доходів за шкалою еквівалентності Європейського Союзу;

19) стратифікація – процес розподілення генеральної сукупності на однорідні групи (страти);

20) територіальна вибірка – відібрані за певною процедурою територіальні одиниці, у яких проводиться вибіркове обстеження;

21) територіальна одиниця – адміністративно-територіальна одиниця України (наприклад: міста, селища міського типу, адміністративні райони, територіальні громади 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).

II. Загальні положення

1. Державні вибіркові обстеження населення є основним джерелом даних для вимірювання багатьох соціально-економічних та демографічних показників, зокрема, вони забезпечують отримання інформації щодо соціально-економічних характеристик домогосподарств, необхідних для макроекономічних розрахунків та розрахунків основних показників у галузі сільського господарства, проведення всебічного аналізу рівня життя населення України, моніторингу бідності, комплексної оцінки стану ринку праці у відповідності до міжнародних стандартів та потреб користувачів інформації всіх рівнів.

2. Основним етапом підготовки та проведення будь-якого вибіркового обстеження є визначення принципів формування вибіркової сукупності респондентів (далі – вибірка). Якість планування вибірки та коректність процедур її формування значною мірою визначають якість результатів вибіркового обстеження, а також адекватність витрат на його проведення.

3. Метою формування вибірки є визначення за спеціально розробленими науково обґрунтованими процедурами частини генеральної сукупності, обстеження якої дозволяє отримати надійну інформацію щодо досліджуваних в обстеженні характеристик генеральної сукупності.

4. Ці Методологічні положення розроблені з метою забезпечення належної якості виконання робіт із планування та формування вибірки з урахуванням наявної інформаційної бази для проведення вибірових обстежень населення.

5. Запровадження вибіркового обстеження EU-SILC на принципах Європейської статистичної практики відповідно до вимог Регламенту (ЄС) № 1177/2003 Європейського Парламенту та Ради від 16 червня 2003 року про статистику Співтовариства щодо доходів і умов життя (EU-SILC) [22] передбачено Програмою розвитку державної статистики до 2023 року. Одним із заходів з реалізації програми є розробка методики формування вибірових сукупностей респондентів у рамках діючої системи проведення вибірових обстежень населення з урахуванням впровадження EU-SILC.

6. Результати обстеження EU-SILC – це основне джерело даних щодо розподілу доходів та умов життя населення в країнах Європейського Союзу. Метою проведення EU-SILC є збір актуальних і зіставних поперечних та поздовжніх (панельних) даних щодо різних аспектів якості життя населення. Результати обстеження EU-SILC використовуються для формування інформаційної бази для аналізу питань бідності та соціальної ізоляції.

7. Відповідно до Регламенту (ЄС) № 1982/2003 [22], відбір одиниць для обстеження EU-SILC здійснюється з урахуванням таких вимог:

поперечні та поздовжні дані мають ґрунтуватися на репрезентативній для країни ймовірнісній вибірці населення, що проживає в приватних домогосподарствах у межах країни, незалежно від мови, національності чи юридичного статусу місця проживання. Для обстеження можуть бути відібраними всі приватні домогосподарства і всі особи, старші за 16 років, що проживають у домогосподарстві;

репрезентативні ймовірнісні вибірки необхідно використовувати як для домогосподарств, які є основними одиницями вибірки та основними одиницями спостереження, так і для осіб - членів домогосподарств;

основа вибірки і методи формування вибірки мають гарантувати, що кожному домогосподарству і кожній особі із сукупності присвоєна певна відома ймовірність відбору, відмінна від нуля.

8. Упровадження обстеження EU-SILC в існуючу систему державних вибірових обстежень населення (домогосподарств) потребуватиме певного уточнення дизайну вибірок діючих обстежень, насамперед у частині обсягів вибірок та їх розподілів за регіонами та типами місцевості. Це зумовлено як необхідністю забезпечення збалансованого загального навантаження фахівців з інтерв'ювання, так і доцільністю оптимізації рівнів надійності оцінювання основних показників у різних обстеженнях.

9. Реалізація заходів реформування місцевого самоврядування та територіальної організації влади України, зокрема щодо утворення та ліквідації районів, також потребуватиме відповідних змін у дизайні вибірок державних вибірових обстежень населення.

III. Організація робіт із формування вибірки

1. Організація робіт із формування вибірових сукупностей населення за умови їх функціонування в рамках системи проведення вибірових обстежень населення складається з таких етапів:

визначення по кожному вибіровому обстеженню вимог до вибірки відповідно до цілей цього державного статистичного спостереження;

формування (актуалізація) основи для кожного етапу вибірки з урахуванням підсумків аналізу якості наявних інформаційних джерел;

розробка плану (дизайну) вибірки для забезпечення функціонування системи вибірових обстежень населення;

розрахунок обсягів вибірових сукупностей по кожному обстеженню;

розробка документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств)" (додаток 1);

реалізація робіт із формування вибірок територій та респондентів.

2. Визначення вимог до вибірки, формування основи вибірки з урахуванням підсумків аналізу якості наявних інформаційних джерел, розробка плану (дизайну) вибірки, розрахунок обсягів вибірових сукупностей, організація робіт, пов'язаних із формуванням вибірки, та розрахунок імовірностей відбору здійснюються на центральному рівні. Роботи із формування основ вибірки та формування вибірових сукупностей територій здійснюються на центральному рівні із залученням до окремих етапів робіт підрозділів територіальних органів Держстату, до функціональних повноважень яких належать питання організації вибірових обстежень населення. Роботи із актуалізації основи вибірки та формування вибірових сукупностей респондентів здійснюються на регіональному рівні.

3. Вимоги до формування вибірки, визначені окремо для кожного з вибірових обстежень населення, формалізуються у вигляді документа, який узгоджується на центральному рівні всіма зацікавленими сторонами та затверджується в установленому порядку в рамках методики щодо формування вибірових сукупностей для проведення органами державної статистики чергового циклу (наприклад п'ятирічного) вибірових обстежень населення.

Необхідність розробки такого комплексного документа зумовлена інтегрованістю вибірових обстежень населення, які на постійній основі проводять органи державної статистики. Зазначені обстеження взаємопов'язані узгодженим планом (дизайном) вибірок, спільною мережею фахівців з інтерв'ювання, узгодженістю графіків їх роботи по кожному з обстежень.

4. Створення документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств)", як правило, має ітераційний характер, оскільки організаційні та фінансові обмеження можуть призвести до неможливості задоволення всіх вимог, висунутих на початку робіт.

5. У випадку необхідності проведення вибірового обстеження населення поза межами системи для цього обстеження також має бути розроблений документ "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств)". На основі цього документа в подальшому приймається рішення щодо проведення окремого обстеження або проведення його у вигляді модуля до функціонуючих обстежень.

IV. Визначення вимог до вибірки відповідно до цілей вибіркового обстеження

1. Відповідно до цілей по кожному вибіркового обстеженню мають бути визначені вимоги до вибірки за такими напрямками:

1) вимоги щодо представлення у вибірці територіальних одиниць відповідно до адміністративно-територіального устрою України.

Визначення цих вимог при плануванні вибірових обстежень населення є обов'язковим.

У випадках обмеженого доступу до окремих населених пунктів або недоцільності з об'єктивних причин проведення в них обстеження можуть бути введені певні обмеження щодо їх уключення до вибірки. Наприклад, до вибірки вибірових обстежень населення недоцільно включати населені пункти, розташовані на територіях радіологічного забруднення, віднесених до зони відчуження (І зона) та зони безумовного (обов'язкового) відселення (ІІ зона) внаслідок катастрофи на Чорнобильській АЕС, також до вибірки не можуть бути включені населені пункти з непідконтрольних Уряду територій АР Крим та Донецької і Луганської областей;

2) вимоги до надійності результатів вибірових обстежень населення.

При плануванні вибірового обстеження слід визначити надійність, з якою необхідно отримати оцінки основних показників для національного рівня та рівня окремих адміністративно-територіальних одиниць. Для кількісної характеристики надійності оцінок показників використовується коефіцієнт варіації вибірових оцінок. Для рівня країни, як правило, оцінки показників вважаються надійними при значеннях коефіцієнта варіації не більше 5%, для регіонального рівня – не більше 10%. У деяких випадках вимоги до надійності можуть установлюватися згідно з положеннями відповідних регламентів ЄС або інших міжнародних нормативних документів.

Від необхідного рівня надійності оцінок основних показників залежить обсяг вибіркової сукупності. При визначенні надійності слід урахувати квадратичну залежність між коефіцієнтом варіації та обсягом вибірки. Наприклад, щоб підвищити надійність (тобто зменшити коефіцієнт варіації) утричі, обсяг вибіркової сукупності необхідно збільшити у дев'ять разів (див. розділ VI цього документа);

3) вимоги до терміну дії територіальної вибірки.

На етапі планування вибірового обстеження визначається строк дії територіальної вибірки з урахуванням її потенціалу для забезпечення проведення визначеної кількості запланованих вибірових обстежень у відібраних територіальних одиницях, а також особливостей організації роботи мережі фахівців з інтерв'ювання. Строк дії територіальної вибірки вказується у роках.

Строк дії територіальної вибірки враховується при визначенні мінімального розміру територіальних одиниць – кількості домогосподарств (осіб), які проживають у них;

4) початкове навантаження фахівця з інтерв'ювання.

Початкове навантаження (кількість одиниць спостереження, які підлягають обстеженню одним фахівцем з інтерв'ювання протягом визначеного на етапі планування проміжку часу) є характеристикою, яка використовується для організації роботи фахівців з інтерв'ювання та формування територіальної вибірки.

При визначенні навантаження слід урахувувати всі компоненти витрат часу фахівця з інтерв'ювання, а саме:

- підготовчі роботи (самопідготовка, підготовка до інтерв'ю);
- проведення опитування в домогосподарствах та збір іншої інформації, передбаченої програмою вибіркового обстеження;
- час, необхідний, щоб дістатись до респондента (транспорт, пішки);
- перевірка, кодування (якщо це передбачено функціональними обов'язками в частині окремих вибірових обстежень) та здавання анкет до територіальних органів Держстату;
- навчання тощо.

Тривалість проведення інтерв'ю (опитування) визначається в пілотному обстеженні або на підставі фактичних витрат часу (якщо обстеження проводилось раніше).

При формуванні територіальної вибірки початкове навантаження фахівця з інтерв'ювання враховується:

- при визначенні порогу саморепрезентативності;
- при визначенні мінімального розміру територіальної одиниці;

5) вимога щодо врахування очікуваного рівня участі.

Для забезпечення обстеження визначеної кількості одиниць спостереження і зменшення впливу відмов респондентів від участі в обстеженні на результати вибіркового обстеження при формуванні вибірки слід урахувувати очікуваний рівень участі. Цей показник використовується для визначення кількості одиниць спостереження, яку необхідно відібрати.

Рівень участі розраховується за загальною формулою:

$$k_{\pi} = cf / c , \quad (1)$$

де k_{π} – рівень участі домогосподарств;

cf – кількість одиниць спостереження (осіб, домогосподарств), які взяли участь в обстеженні;

c – кількість одиниць спостереження, які були відібрані для участі в обстеженні.

Очікуваний рівень участі визначається за даними аналогічних попередніх вибірових обстежень (за їх наявністю), за результатами пілотних обстежень або

за експертними оцінками. Очікуваний рівень участі використовується при визначенні обсягу вибіркової сукупності та мінімального розміру територіальної одиниці.

Якщо за результатами вибіркового обстеження збираються дані як щодо домогосподарств, так й щодо їх членів (це характерно для обстежень ОУЖД, ОРС, EU-SILC), доцільно окремо враховувати рівні участі домогосподарств та осіб;

6) схема ротації одиниць вибірки.

Для підвищення якості результатів вибірових обстежень населення, що проводяться на постійній основі, застосовується спеціальна процедура – повна або часткова ротація одиниць вибірки, тобто їх планова заміна за спеціально визначеними схемами. Повна ротація передбачає одночасну заміну всіх одиниць вибірки, часткова – заміну частини вибірки відповідно до схеми ротації. Ротація одиниць вибірки забезпечує покращення якості даних за рахунок зниження навантаження на респондентів, оскільки останні перебувають в обстеженні визначений схемами час (певну кількість періодів обстеження). Крім того, часткова ротація респондентів забезпечує більш надійне вимірювання зміни показників. Застосування цієї процедури регламентується методологією кожного з обстежень.

Для здійснення часткової ротації вибіркова сукупність розбивається на групи одиниць вибірки – ротаційні групи. Ротаційні групи відбираються як окремі незалежні підвибірки. Схема ротації одиниць вибірки регламентується методологією певного обстеження.

Приклад схеми ротації домогосподарств при проведенні у 2019–2023 роках обстеження робочої сили наведено в додатку 2.

Іншим прикладом ротаційного дизайну вибірки є дизайн вибірки в обстеженні EU-SILC, а саме: вибірка формується на основі низки підвибірок (найчастіше чотирьох), ідентичних за дизайном та близьких за обсягом. З року в рік одні підвибірки зберігаються, інші відкидаються і замінюються новими підвибірками. У разі ротаційного дизайну на основі чотирьох підвибірок з ротацією однієї підвибірки на рік одна з підвибірок виключається відразу ж після першого року, друга має зберігатися протягом двох років, третя протягом трьох років, четверта повинна зберігатися протягом чотирьох років. З другого року і надалі одна нова підвибірка вводиться щороку і зберігається протягом чотирьох років. Приклад ротаційного дизайну вибірки, який рекомендується для України при впровадженні обстеження EU-SILC, наведено у додатку 3.

Застосування часткової ротації в цілому збільшує кількість обстежених одиниць вибірки за певний період. У зв'язку з цим ротацію слід урахувати при визначенні кількості одиниць вибірки, яку необхідно відібрати на рік, а також мінімального розміру територіальної одиниці;

7) додаткові вимоги.

Додаткові вимоги до формування вибірки передбачають забезпечення інших особливостей відповідно до програм вибірових обстежень населення або

забезпечення функціонування їх системи. Ці вимоги враховуються при плануванні та реалізації етапів формування вибірки.

У випадку формування вибірки для проведення декількох обстежень може бути визначена вимога щодо обмеження участі в обстеженнях одних і тих самих одиниць вибірки. Наприклад, якщо два обстеження планується проводити в одній територіальній одиниці і визначене обмеження участі в обстеженнях одних і тих самих одиниць вибірки, то мінімальний розмір територіальної одиниці розраховується як сума мінімальних розмірів, необхідних для проведення кожного з цих обстежень.

2. У випадку формування вибірки для проведення декількох обстежень окремо визначаються вимоги до формування вибіркової сукупності кожного обстеження з подальшим їх взаємоузгодженням.

V. Формування основи вибірки з урахуванням підсумків аналізу якості наявних інформаційних джерел

1. Визначальною умовою розробки ефективного плану (дизайну) вибірки є формування якісної основи вибірки. Це передбачає дослідження та вибір найбільш придатних інформаційних джерел, які мають містити інформацію щодо територіального розміщення одиниць відбору та характеристики цих одиниць, необхідні для формування вибіркової сукупності. Інформація, що використовується при формуванні основи вибірки, має задовольняти таким вимогам: повнота, якість, актуальність.

Вимога щодо повноти інформації означає, що необхідні для відбору характеристики є в наявності у визначеному джерелі для всіх без винятку одиниць генеральної сукупності.

Вимога щодо якості означає, що дані, наявні у визначеному інформаційному джерелі, відповідають реальному (актуальному) стану генеральної сукупності, отримані з офіційних джерел та користуються довірою фахівців.

Вимога щодо актуальності означає, що інформація, яка використовується при формуванні основи вибірки, за часом повинна бути максимально наближена до часу формування вибірки.

2. Для формування основи вибірки з метою здійснення певного ступеня відбору повинна використовуватись інформація з одного джерела. Для формування основи багатоступеневої вибірки на різних ступенях відбору може використовуватись інформація з різних джерел.

Основа вибірки для проведення вибірових обстежень населення потенційно може формуватися на основі використання даних з різних джерел, зокрема:

переписів населення (щодо розміщення та кількості постійного населення по кожній адміністративно-територіальній одиниці, кількості та розміру лічильних ділянок тощо);

демографічної статистики (щодо розміщення та кількості постійного та наявного населення по окремих адміністративно-територіальних одиницях);

сільськогосподарських переписів;
 погосподарського обліку;
 картографічних матеріалів;
 матеріалів проведення виборчих кампаній (щодо розміщення та характеристики виборчих діляниць, утворених на постійній основі на території окремих населених пунктів);
 поділу території країни на поштові відділення та кількості населення, що обслуговується ними;
 баз даних окремих міністерств і відомств з інформацією про все населення або окремі його групи (Єдиний державний демографічний реєстр, Державний реєстр фізичних осіб – платників податків, Реєстр застрахованих осіб Пенсійного фонду України тощо).

3. Формування щорічних вибірових сукупностей для вибірових обстежень населення (домогосподарств) здійснюється з використанням основи вибірки, інформаційна база якої обов'язково повинна бути актуалізована безпосередньо перед початком процедур відбору.

4. На окремих етапах формування вибіркової сукупності може виникати необхідність використання персоніфікованої інформації щодо одиниць генеральної сукупності. Забезпечення статистичної конфіденційності здійснюється відповідно до вимог статей 14, 16, 21, 22 Закону України "Про державну статистику" [1], статей 18, 21 Закону України "Про інформацію" [2], статей 6 і 7 Закону України "Про доступ до публічної інформації" [3], статей 5, 6, 7, 24 Закону України "Про захист персональних даних" [4] з урахуванням основних принципів і методів, визначених Методологічними положеннями щодо забезпечення статистичної конфіденційності в органах державної статистики [5], міжнародними вимогами до правил конфіденційності статистичної інформації відповідно до статей 20–26 Регламенту Європейського Парламенту та Ради від 11 березня 2009 року № 223/2009 "Про європейську статистику" [24]. Для забезпечення встановлених законодавством гарантій перед респондентами щодо забезпечення конфіденційності наданої ними статистичної інформації використання цієї інформації відбувається виключно у статистичних цілях.

VI. Розрахунок обсягу вибіркової сукупності

1. Одним з визначальних етапів планування вибіркового обстеження та формування вибірки є визначення обсягу вибіркової сукупності, необхідного для задоволення встановлених вимог. Від обсягу вибірки залежать як якість результатів обстеження, так і обсяг витрат ресурсів на його проведення. Занадто великий обсяг вибірки призводить до невиправданих витрат та неефективного використання коштів, а недостатній обсяг вибірки – до незадовільної якості результатів обстеження і відповідно також до невиправданості витрат на проведення обстеження.

2. Оцінка обсягу вибірки здійснюється з урахуванням таких основних факторів [18]:

вимог щодо надійності оцінок основних показників, перелік яких має бути визначений перед початком робіт із планування вибірки. Кількість таких показників, як правило, становить від 1 до 3;

обсягу фінансових коштів, виділених для проведення обстеження;

кількості фахівців з інтерв'ювання, які будуть проводити обстеження, та їх навантаження (кількості респондентів, яких вони мають опитати протягом певного періоду).

У випадку формування вибірки для проведення декількох обстежень оцінка обсягу вибірки здійснюється окремо по кожному з них.

3. Досягнутий обсяг стратифікованої багатоступеневої вибірки у випадках необхідності отримання надійних оцінок середніх значень розраховується за формулою:

$$n_d = deff_{\bar{y}} \times \sigma^2 \times \left(\frac{100\%}{CV_{\bar{y}} \times \bar{y}} \right)^2 \quad (2)$$

де n_d – досягнутий обсяг вибірки;

σ^2 – дисперсія, яка характеризує варіацію значень ознаки, за якою розраховується показник, по одиницях генеральної сукупності;

$\bar{y}$ – оцінка очікуваного середнього значення показника (наприклад, середньомісячні сукупні ресурси домогосподарства на одну особу);

$deff_{\bar{y}}$ – значення дизайн-ефекту при оцінюванні середнього значення $\bar{y}$.

Дизайн-ефект є відносним показником, який кількісно характеризує вплив плану (дизайну) вибірки на надійність оцінок показників вибірових обстежень порівняно з вибіркою, побудованою за методом простого випадкового відбору. Для простої випадкової вибірки $deff_{\bar{y}} = 1$;

$CV_{\bar{y}}$ – коефіцієнт варіації вибірових оцінок середнього значення $\bar{y}$ для показника, визначений у вимогах до формування вибірки, який розраховується за формулою:

$$CV_{\bar{y}} = \frac{SE_{\bar{y}}}{\bar{y}} \times 100\%, \quad (3)$$

де $SE_{\bar{y}}$ – стандартна похибка оцінювання середнього значення $\bar{y}$, яка розраховується на підставі величини оцінки дисперсії σ^2 та обсягу вибірки n за формулою:

$$SE_{\bar{y}} = \sqrt{\frac{\sigma^2}{n}}. \quad (4)$$

Величина σ^2 може бути оцінена за результатами вибірового обстеження за формулою:

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}, \quad (5)$$

де y_i – значення ознаки для i -го домогосподарства (наприклад, якщо y – грошовий дохід, то y_i – грошовий дохід певного (i -го) домогосподарства з вибірки).

Значення коефіцієнта варіації встановлюється відповідно до вимог щодо надійності оцінки показника. Величини дизайн-ефектів, дисперсій та оцінок показників визначаються за даними аналогічних попередніх вибірових обстежень населення (за наявності таких даних), за результатами пілотних обстежень або за експертними оцінками. Наприклад, за рекомендаціями міжнародних експертів, які надавали консультаційну допомогу органам державної статистики з питань упровадження ОУЖД та ОРС, значення дизайн-ефекту при оцінюванні середніх значень доходів та витрат домогосподарств на національному рівні приймалося рівним 2,5 з урахуванням досвіду проведення аналогічних обстежень в інших країнах та наявного досвіду проведення цих обстежень в Україні.

4. У випадках необхідності отримання надійних оцінок для показників часток та сумарних значень обсяги вибірок зі складним дизайном розраховуються за формулами, аналогічними формулі (2).

При оцінюванні часток (p) формула для розрахунку обсягу вибірки набуває вигляду:

$$n = def f_p \times p \times (1 - p) \times \left(\frac{100\%}{CV_q \times p} \right)^2 = def f_p \times \frac{(1-p)}{p} \times \left(\frac{100\%}{CV_p} \right)^2. \quad (6)$$

При цьому враховано, що дисперсія σ^2 для часток визначається за формулою [9]:

$$\sigma^2 = p \times (1 - p) \quad (7)$$

де p – оцінка очікуваного значення частки (наприклад, частки населення із середньодушовими еквівалентними загальними доходами у місяць нижче середньорічного фактичного прожиткового мінімуму).

При оцінюванні сумарних значень формула для розрахунку обсягу вибірки набуває вигляду:

$$n = def f_{\hat{Y}} \times \sigma^2 \times N^2 \times \left(\frac{100\%}{CV_{\hat{Y}} \times \hat{Y}} \right)^2, \quad (8)$$

де N – обсяг генеральної сукупності;

$\hat{Y}$ – оцінка очікуваного сумарного значення показника (наприклад, площі землі домогосподарств). При цьому враховано, що $\hat{Y} \approx \bar{y} \times N$.

5. У деяких випадках нормативні документи можуть визначати правила розрахунку обсягу вибірки. Зокрема, підходи до визначення ефективного обсягу вибірки в обстеженні EU-SILC визначаються Регламентом (ЄС) № 2019/1700.

У відповідності до цього регламенту при визначенні обсягу вибірки домогосподарств в обстеженні EU-SILC вимоги до надійності виражаються через

обмеження величини стандартної похибки оцінювання показників, яка визначається як функції оцінок показників та обсягу генеральної сукупності в країні за формулою (9). Передбачається, що стандартна похибка $\widehat{SE}(\hat{p})$ оцінки показника $\hat{p}$ не має перевищувати певну величину:

$$\widehat{SE}(\hat{p}) \leq \sqrt{\frac{\hat{p}(1-\hat{p})}{f(N)}}, \quad (9)$$

де $f(N) = a\sqrt{N} + b$.

При цьому для параметрів N , a і b використовуються рекомендовані значення (див. табл. 1).

Таблиця 1

Параметри N , a і b для визначення обсягу вибірки за показниками доходів та умов життя в обстеженні EU-SILC (Регламент ЄС № 2019/1700, Додаток II)

$\hat{p}$	N	a	b
Ризик бідності або соціального відчуження на національному рівні	Кількість приватних домогосподарств у країні в мільйонах і округлюється до 3 знаків після коми	900	2600
Рівень стійкого ризику бідності протягом чотирьох років на національному рівні	Кількість приватних домогосподарств у країні в мільйонах і округлюється до 3 знаків після коми	350	1000
Ризик бідності або соціального відчуження на регіональному рівні	Кількість приватних домогосподарств у регіоні NUTS 2 у мільйонах та округлена до трьох знаків після коми	600	0

Для забезпечення умови (9) обсяг ефективної вибірки на національному рівні має задовольняти такому співвідношенню:

$$n_{eff} \geq \frac{\hat{p}(1-\hat{p})}{\widehat{SE}(\hat{p})^2} = f(N). \quad (10)$$

Для оцінки ризику бідності або соціального відчуження для населення вимоги, визначені за формулою (9), не є обов'язковими для регіонів, у яких проживає менше 500 тисяч осіб.

На основі зазначеного підходу визначено мінімальні ефективні обсяги вибірки EU-SILC для країн ЄС (додаток 4).

Слід урахувати, що ефективний обсяг вибірки є обсягом, розрахованим для обстеження, у якому реалізується простий випадковий відбір одиниць спостереження (домогосподарств або осіб). Дизайн-ефект для такої вибірки дорівнює 1,0. Фактичний обсяг вибірки має бути більшим ефективного обсягу настільки, щоб компенсувати вплив реального дизайн-ефекту, який, як правило, перевищує 1,0 і в обстеженнях домогосподарств може досягати 2,5 і більше, а

також щоб компенсувати ефект відмов відібраних одиниць від участі у спостереженні. Крім того, слід ураховувати, що ефективний обсяг вибірки складається лише з кількості обстежених домогосподарств (осіб), тобто лише з тих одиниць спостереження, по яких одержано інформацію необхідної повноти.

6. При визначенні у вимогах декількох показників, для яких необхідно отримати оцінки встановленого рівня надійності, обсяг вибіркової сукупності розраховується для кожного з них окремо. Максимальне з отриманих значень має бути прийняте за підсумковий обсяг вибірки, оскільки при цьому забезпечується необхідний рівень надійності всіх основних показників.

7. У випадку визначення у вимогах необхідності отримання надійних оцінок показників для окремих частин генеральної сукупності, наприклад для регіонів, обсяг вибірки розраховується по кожній частині сукупності окремо. Сума обсягів вибірок по всіх частинах генеральної сукупності приймається за підсумковий обсяг вибірки по Україні.

Обсяг вибіркової сукупності коригується, якщо визначена вимога врахування очікуваного рівня участі.

8. Визначення необхідного досягнутого обсягу вибірки n_d за відомого ефективного обсягу n_{eff} вибірки здійснюється з урахуванням дизайн-ефекту $deff$ за формулою:

$$n_d = deff \times n_{eff}. \quad (11)$$

9. Розрахунок обсягу вибірки (див., наприклад, формулу (2)) з урахуванням очікуваного рівня участі домогосподарств (k_{rr}) здійснюється за такою формулою:

$$n_{факт} = \frac{n}{k_{rr}}, \quad (12)$$

де $n_{факт}$ – початковий (фактичний) обсяг вибірки з урахуванням очікуваного рівня участі домогосподарств.

10. За умови, що обстеження проводиться на постійній основі (щомісячно, щоквартально, щорічно) і план (дизайн) вибірки не зазнає суттєвих змін, фактичний (діючий) обсяг вибірки такого обстеження може бути скоригований на основі досвіду проведення обстежень у попередніх циклах, стану явищ та процесів, що оцінюються, вимог до проведення обстежень. Наприклад, при визначенні загального обсягу вибірки на наступний цикл за обсягами вибірки для регіонів розрахунки здійснюються за формулою:

$$n_{вимог} = \sum_j (n_{факт,j} \times \frac{CV_{факт,j}^2}{CV_{вимог}^2}), \quad (13)$$

де $n_{\text{вимог}}$ – теоретичний загальний обсяг вибірки з урахуванням вимог щодо рівня надійності цільового показника на регіональному рівні;

$n_{\text{факт},j}$ – фактичний обсяг вибірки для j -го регіону за результатами обстеження за останній рік попереднього циклу;

$CV_{\text{факт},j}$ – середнє фактичне значення коефіцієнта варіації для показника за три останні роки попереднього циклу проведення обстеження для j -го регіону (використання середнього значення коефіцієнта варіації забезпечує більшу надійність його оцінки);

$CV_{\text{вимог}}$ – значення коефіцієнта варіації, зазначене у вимогах для регіонального рівня ($CV_{\text{вимог}} = 10\%$).

11. При розрахунках обсягу вибірки для проведення вибірових обстежень населення може застосовуватись процедура оптимізації дизайну вибірки з метою підвищення надійності оцінювання певних визначених показників на регіональному рівні.

Особливості здійснення процедури оптимізації визначаються окремими методичними документами для кожного вибірового обстеження населення.

12. Для ефективного розподілу визначених обсягів вибірок за стратами (наприклад: "великі міста регіону", "малі міста регіону" та "райони регіону"), як правило, використовується формула оптимального розміщення Неймана [12]:

$$n_{j,l} = n_j \times \frac{W_{j,l} \times \sigma_{j,l}}{\sum_l W_{j,l} \times \sigma_{j,l}} \quad (14)$$

де $n_{j,l}$ – обсяг вибірки для l -ї страти j -го регіону ($l=1,2,3$);

n_j – обсяг вибірки по певному обстеженню для j -го регіону;

$W_{j,l}$ – частка загальної кількості постійного населення регіону, що припадає на l -ту страту j -го регіону;

$\sigma_{j,l}$ – оцінка середньоквадратичного відхилення значень показника l -ї страти j -го регіону.

Такий розподіл обсягу вибірки за стратами забезпечує максимальну надійність оцінювання показника на рівні регіону за умови, що загальний обсяг вибірки для регіону є визначеним.

13. Розрахований для встановленого рівня надійності оцінок показників обсяг вибіркової сукупності порівнюється із фінансовими та інституційними можливостями органів державної статистики щодо забезпечення проведення вибірового обстеження. З цією метою визначається необхідна кількість фахівців з інтерв'ювання, які мають проводити польовий етап обстеження, інших працівників, які будуть забезпечувати розробку й аналіз результатів, а також розраховується необхідний обсяг фінансових витрат. У випадку недостатності

можливостей та коштів для забезпечення проведення обстеження за визначених умов необхідно здійснити коригування вимог, наприклад, у напрямі зменшення надійності оцінок показників. Після цього розраховується новий обсяг вибіркової сукупності і процедура порівняння з можливостями здійснюється повторно. Таким чином, процедура визначення обсягу вибірки є ітераційною, яка завершується після узгодження всіх вимог до вибірки з фінансовими та інституційними можливостями органів державної статистики.

VII. Розробка документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств)"

1. Документ "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств)" (далі – Вимоги) узагальнює підсумки попередніх етапів визначення та узгодження всіх вимог до вибірки та впливає на подальшу організацію робіт з побудови вибіркової сукупності, проведення обстежень та поширення їх результатів на генеральну сукупність.

Вимоги до формування вибірки офіційно узгоджуються з усіма зацікавленими самостійними структурними підрозділами Держстату.

2. Документ "Вимоги" включає такі основні розділи:

1) мета побудови вибірки;

2) генеральна сукупність;

3) методологія побудови вибірки;

4) одиниці спостереження та одиниці відбору. При визначенні одиниць спостереження та одиниць відбору слід мати на увазі, що можливі випадки, коли ці одиниці не збігаються. Наприклад, при визначенні домогосподарства як одиниці відбору одиницею спостереження відповідно до мети обстеження може бути окрема особа, яка входить до складу домогосподарства;

5) вимоги щодо представлення у вибірці територіальних одиниць відповідно до адміністративно-територіального устрою України. Наприклад, необхідність забезпечення представлення у вибірці всіх регіонів України, а також за необхідністю – інших адміністративно-територіальних одиниць: обласних центрів, районів, об'єднаних місцевих територіальних громад, районів у містах тощо;

6) термін дії територіальної вибірки;

7) тип вибірки, метод та одиниці відбору. Наприклад, необхідність застосування багатоступеневої вибірки з визначенням способу відбору на кожному ступені, зокрема систематичного відбору, з імовірністю, пропорційною розміру, простого випадкового відбору та ін.;

8) кількість територіальних одиниць для відбору для проведення в них вибіркового обстеження одним фахівцем з інтерв'ювання у міській та сільській місцевостях;

9) обмеження щодо представлення у вибірці окремих груп одиниць спостереження (наприклад, в EU-SILC: "Особи, які проживають в колективних домогосподарствах і установах, виключаються з цільової сукупності" [21]);

10) перелік основних показників, надійність яких має бути забезпечена у вибіркому спостереженні та рівні надійності оцінювання цих показників для національного рівня, а також за необхідності – у розрізі окремих соціально-економічних груп або для різних територіальних рівнів;

11) навантаження фахівця з інтерв'ювання;

12) очікуваний рівень участі;

13) термін дії вибірки респондентів;

14) наявність і схема ротації одиниць спостереження;

15) обсяг вибірки респондентів з урахуванням визначених вимог (може вказуватися період дії вибірки, наприклад: рік; квартал; місяць);

16) додаткові вимоги до формування вибірки (наприклад, ОУЖД має бути оптимізована для визначених показників бідності; можуть зазначатись вікові обмеження, зокрема, в ОРС опитуванню підлягають усі особи віком 15 років і старше, у EU-SILC – 16 років і старше [21] тощо).

3. У випадку формування вибірки для одночасного проведення декількох вибірових обстежень населення вищенаведені положення наводяться для всіх, але визначаються для кожного з обстежень окремо.

4. Ураховуючи особливості проведення опитувань населення, вимоги щодо початкового навантаження одного фахівця з інтерв'ювання та очікуваного рівня участі респондентів визначаються в розрізі типів місцевості проживання домогосподарств.

5. У разі проведення фахівцем з інтерв'ювання декількох вибірових обстежень одночасно, але на різних мережах респондентів, під час узгодження

вимог початкове навантаження коригується з урахуванням законодавчо встановленої тривалості робочого дня.

6. Макет документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірових обстежень населення (домогосподарств)" наведений у додатку 1.

7. Приклад матеріалів та даних для формування документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибірового обстеження EU-SILC" представлено у додатку 5.

VIII. Розробка плану (дизайну) вибірки

1. Після затвердження Вимог здійснюється розробка плану (дизайну) вибірки – організаційно-логічної моделі структури вибірової сукупності та принципів її формування. На цьому етапі також опрацьовуються зміст і послідовність усіх процедур формування вибірки. Результатом цієї роботи має бути методика формування вибірових сукупностей для проведення вибірових обстежень населення (домогосподарств) протягом певного календарного циклу.

2. При розробці плану (дизайну) вибірки необхідно враховувати обмеження та умови, які визначаються Вимогами.

План (дизайн) вибірки має забезпечувати формування ймовірнісної вибірки і враховувати Вимоги щодо формування вибірки.

Далі наведені основні методологічні підходи до розробки плану (дизайну) територіальних вибірок для вибірових обстежень населення:

визначення обмежень щодо охоплення вибіровим обстеженням адміністративно-територіальних одиниць України;

стратифікація генеральної сукупності;

визначення типу вибірки і способів відбору на кожному етапі.

3. Визначення обмежень щодо охоплення вибіровим обстеженням адміністративно-територіальних одиниць України.

На початку формування територіальної вибірки визначаються адміністративно-територіальні одиниці України, які не повинні або не можуть бути обстежені з об'єктивних причин. Обмеження щодо територіального охоплення відображаються у Вимогах та враховуються при реалізації процедур відбору й обробки результатів обстеження.

4. Стратифікація генеральної сукупності.

Формування вибірових сукупностей для проведення вибірових обстежень населення передбачає застосування на початковому етапі відбору стратифікації генеральної сукупності. Стратифікація здійснюється з метою адекватного відображення у вибірці основних особливостей адміністративно-територіального

поділу країни, а також для забезпечення відбору домогосподарств із більш однорідних за основними характеристиками груп [6, 7, 11, 17, 18].

Для реалізації цієї мети генеральна сукупність розподіляється на страти.

Наприклад, в межах кожного регіону страти можуть відповідати:

містам і міськрадам із кількістю населення 100 тис. осіб і більше (далі – страта "великі міста регіону");

містам, селищам міського типу, міські територіальні громади із кількістю населення менше 100 тис. осіб (далі – страта "малі міста регіону");

адміністративним районам сільської місцевості (далі – страти "райони регіону").

Такий розподіл забезпечує потреби щодо отримання інформації за типами місцевості (міська, сільська), а також надає можливість урахувати різні умови роботи фахівців з інтерв'ювання у різних за розміром та за типами населених пунктах (різні витрати часу на проїзд, різне навантаження), різні інформаційні джерела, необхідні для відбору територіальних одиниць та одиниць спостереження тощо.

Обсяг вибірки розподіляється між стратами за певними принципами, наприклад, пропорційно розміру страт (кількості населення або домогосподарств в них) або за формулою Неймана.

Принципи розподілу обсягу вибірки за стратами визначаються в залежності від завдань забезпечення ефективного вимірювання значної кількості показників вибірових обстежень на національному або регіональному рівні.

5. Визначення типу вибірки:

1) у відповідності до встановлених вимог до вибірки визначаються особливості її формування: основа вибірки, кількість ступенів відбору, способи відбору на кожному ступені та ін.

За типом вибірка може бути проступеневою або багаступеневою.

Формування проступеневої вибірки здійснюється в один етап – шляхом безпосереднього відбору одиниць генеральної сукупності. При вибірових обстеженнях населення, які проводяться шляхом відвідування домогосподарств, як правило, недоцільно використовувати проступеневі вибірки. Це пов'язано з тим, що такий тип вибірки:

потребує основи вибірки, яка містила б перелік усіх одиниць генеральної сукупності (наприклад, повний перелік домогосподарств (населення) із зазначенням їх адрес);

призводить до значного розшарування одиниць спостереження по території, що обстежується, та збільшення витрат часу та фінансових ресурсів, необхідних для проведення обстеження.

У випадку застосування стратифікації генеральної сукупності формування вибірки проводиться всередині кожної страти.

Формування багаступеневої вибірки здійснюється у кілька послідовних етапів, кількість яких залежить від наявної інформаційної бази, організаційних можливостей органів державної статистики та трудомісткості робіт із формування

вибірки. На різних ступенях відбору можуть визначатися різні одиниці відбору, а також використовуватись різні інформаційні джерела;

2) визначаються одиниці відбору на кожному ступені формування вибірки:

об'єкти адміністративно-територіального устрою країни (наприклад, області, міста, селища міського типу, адміністративні райони, територіальні громади, окремі села);

локальні території, визначені за певним принципом (наприклад, виборчі ділянки; переписні відділи, інструкторські та лічильні ділянки, створені для проведення перепису населення; території, які обслуговуються одним поштовим відділенням, тощо);

домогосподарства;

особи.

Визначення одиниці відбору на конкретному етапі залежить від наявності інформації, необхідної для проведення відбору. У зв'язку з цим на початку робіт слід визначити тип локальних територій, по яких можливо отримати основу вибірки для проведення одного чи декількох етапів відбору, а на останньому етапі – провести безпосередньо відбір одиниць спостереження.

При розробці етапів формування вибірки визначають мінімально необхідну кількість ступенів відбору, що сприяє:

скороченню обсягів робіт щодо формування вибіркової сукупності;

зменшенню похибки вибірки, оскільки кожен додатковий етап відбору, як правило, впливає на похибку вибірки у бік її збільшення.

На вибір типу вибірки також впливає організаційна структура мережі фахівців з інтерв'ювання. Так, особливістю організації мережі фахівців з інтерв'ювання в органах державної статистики є їх проживання у межах тих міст, у яких вони проводять вибірові обстеження населення;

3) результати аналізу наявних інформаційних джерел, вимог до формування вибірок для проведення вибірових обстежень населення (домогосподарств) та організації роботи мережі фахівців з інтерв'ювання зумовили необхідність застосування багатоступеневої вибірки. Наприклад, в ОУЖД, ОРС та EU-SILC вибірка має бути триступеневою, а в ОСГД – двоступеневою. Формування триступеневої вибірки включає такі основні етапи:

відбір ПТОВ – адміністративно-територіальних одиниць;

відбір ВТОВ – локальних територій;

відбір одиниць третього ступеня – домогосподарств (або осіб).

Метою відбору ПТОВ є отримання переліку адміністративно-територіальних одиниць (міських (селищних) рад, міст, селищ міського типу, сільських населених пунктів), у яких будуть проводитись вибірові обстеження населення. Метою відбору ВТОВ є отримання переліку локальних територій у межах ПТОВ (інструкторських ділянок у міській місцевості тощо), у яких на третьому етапі будуть відбиратися одиниці третього ступеня.

Доцільність формування та відбору ВТОВ зумовлена необхідністю оптимізації роботи фахівців з інтерв'ювання шляхом формування та відбору в

межах одиниць першого ступеня, які мають велику за розміром площу, менших локальних територій, по яких існує необхідна для відбору інформація щодо населення;

4) на різних ступенях відбору, відповідно до визначених одиниць відбору та наявних інформаційних джерел, можуть бути різні основи вибірки. Наприклад, при відборі ПТОВ і ВТОВ основу вибірки можуть становити переліки відповідних територіальних одиниць та їх окремі ознаки (наприклад, код відповідно до Кодифікатора адміністративно-територіальних одиниць та територій територіальних громад та кількість населення або кількість домогосподарств у цих одиницях); при відборі домогосподарств основою вибірки є повний список їх адрес по відповідній територіальній одиниці;

5) в окремих випадках, коли деякі територіальні одиниці складаються з певної кількості населених пунктів, припустимим є поділ таких територіальних одиниць на окремі частини (групи населених пунктів) з подальшим їх послідовним обстеженням. Ця процедура фактично є додатковим ступенем відбору, який реалізується в межах окремих територіальних одиниць. Наприклад, об'єднані територіальні громади можуть включати значну кількість сільських населених пунктів, транспортне сполучення між якими є ускладненим. У цьому випадку, з метою забезпечення ефективної роботи фахівця з інтерв'ювання, формуються групи сільських населених пунктів з урахуванням їх територіального розташування. Після цього визначається послідовність (випадковим чином) обстеження груп протягом строку дії територіальної вибірки. Відбір домогосподарств здійснюється в межах кожної сформованої групи сільських населених пунктів.

6. Визначення методів відбору.

На різних ступенях відбору, відповідно до цілей вибіркового обстеження та наявних інформаційних джерел, можуть бути різні методи відбору. Дизайн вибірки і методи формування вибірки повинні забезпечувати, щоб кожній особі (домогосподарству) у цільовій сукупності присвоювалась відома ненульова ймовірність вибору [21].

Основні методи здійснення відбору такі:

простий випадковий;

систематичний;

систематичний з ймовірністю, пропорційною розміру.

Одиниці вибірки можуть мати однакові ймовірності відбору або різні.

Для вибірових обстежень населення одиниці вибірки територій мають різні ймовірності відбору, які є пропорційними розміру територіальної одиниці. Такі ймовірності є результатом застосування відбору з ймовірністю, пропорційною розміру (ВІР) (probability proportional to size, скорочено – PPS). Як розмір може виступати кількість населення або кількість домогосподарств територіальної одиниці, розмір площі землі домогосподарств тощо.

Основна перевага методу відбору з імовірністю, пропорційною розміру, полягає в тому, що для багатоступеневих вибірок він дозволяє забезпечити однакову загальну ймовірність відбору одиниць вибірки останнього ступеня за умови їх рівномірного розподілу по територіальних одиницях різного розміру.

На окремих етапах формування вибірки після упорядкування одиниць відбору за певним визначеним принципом безпосередньо відбір здійснюється із застосуванням систематичного відбору.

ІХ. Реалізація робіт із формування вибірки в рамках системи державних вибірових обстежень населення (домогосподарств)

1. Відбір первинних територіальних одиниць вибірки

1. Реалізація робіт із формування вибірки здійснюється у кілька послідовних етапів із застосуванням процедур, визначених при розробці плану (дизайну) вибірки. Зокрема, формування вибірки для вибірових обстежень населення може здійснюватися у три етапи, а саме: відбір первинних територіальних одиниць вибірки, відбір вторинних територіальних одиниць вибірки, відбір домогосподарств, осіб.

2. Етап відбору первинних територіальних одиниць вибірки складається з таких робіт: визначення порога саморепрезентативності територіальних одиниць, визначення мінімального розміру територіальної одиниці, відбір несамопрезентативних територіальних одиниць, розрахунок імовірностей відбору первинних територіальних одиниць:

1) визначення порога саморепрезентативності.

При проведенні вибірового обстеження, з одного боку, кожна одиниця вибірової сукупності репрезентує певну кількість елементів генеральної сукупності. З іншого боку, фахівець з інтерв'ювання має певне навантаження щодо кількості одиниць вибірової сукупності, яку він повинен обстежити. Таким чином, визначена кількість одиниць вибірової сукупності, яку опитує один фахівець з інтерв'ювання, репрезентує визначену кількість елементів генеральної сукупності. Ця кількість елементів генеральної сукупності називається порогом саморепрезентативності, який у разі застосування методу відбору PPS є критерієм визначення територіальних одиниць, що обов'язково мають бути відібрані до територіальної вибірки, тобто включені до територіальної вибірки з імовірністю, що дорівнює одиниці. Такі територіальні одиниці називаються саморепрезентативними. Несамопрезентативні територіальні одиниці відбираються із застосуванням методу систематичного відбору з імовірністю, пропорційною розміру, та включаються до територіальної вибірки з імовірністю, меншою одиниці.

При формуванні територіальної вибірки для проведення вибірових обстежень населення кількість територіальних одиниць, яку необхідно відібрати, невідома. У зв'язку з цим поріг саморепрезентативності визначають на основі

обсягу вибірки n та встановленої кількості домогосподарств, що обстежуються одним фахівцем з інтерв'ювання (навантаження фахівця з інтерв'ювання), за умови повного завантаження одного фахівця з інтерв'ювання при обстеженні цієї територіальної одиниці. Поріг саморепрезентативності визначається за такою формулою:

$$n_p^{HH} = \frac{N}{n} \times m, \quad (15)$$

де n_p^{HH} – поріг саморепрезентативності (кількість домогосподарств);

N – обсяг генеральної сукупності;

n – обсяг вибіркової сукупності;

m – навантаження фахівця з інтерв'ювання (кількість домогосподарств).

Обсяг генеральної сукупності, обсяг вибірки та навантаження фахівця з інтерв'ювання можуть визначатися як кількістю домогосподарств, так і кількістю населення. Розмір територіальних одиниць (міст, смт) часто визначається в особах. У цьому випадку для розрахунку порога саморепрезентативності (кількість населення) у формулу 15 слід увести додатковий множник – середній розмір домогосподарства:

$$n_p^{pers} = \frac{N}{n} \times m \times S, \quad (16)$$

де n_p^{pers} – поріг саморепрезентативності (кількість населення (осіб));

s – середній розмір домогосподарства (кількість осіб).

Середній розмір домогосподарства у міських населених пунктах може бути отриманий з підсумків перепису населення або з даних попередніх вибірових обстежень населення, а по сільській місцевості – з вищезгаданих джерел, а також із даних форми погосподарського обліку № 1 "Облікова картка об'єкта погосподарського обліку". Доцільність застосування у розрахунках порога саморепрезентативності показника кількості населення пояснюється наявністю даних демографічної статистики щодо кількості населення практично по всіх адміністративно-територіальних одиницях, що дозволяє застосовувати останні для вдосконалення процедур розповсюдження результатів вибірових обстежень населення на генеральну сукупність.

Приклад розрахунку порога саморепрезентативності наведено у додатку 6.

У випадку використання однієї територіальної вибірки для проведення декількох вибірових обстежень населення розрахунок порога саморепрезентативності слід проводити для вибірового обстеження з найбільшим навантаженням фахівця з інтерв'ювання;

2) визначення мінімального розміра територіальної одиниці.

За наявністю вимоги щодо формування вибірових сукупностей для проведення вибірових обстежень населення на певній територіальній одиниці протягом кількох років слід визначити мінімальний розмір такої територіальної одиниці.

Мінімальний розмір територіальної одиниці може бути розрахований за двома підходами в залежності від схеми ротації одиниць вибірки: повної щорічної ротації або часткової.

У випадку повної щорічної ротації одиниць вибірки розрахунок базується на навантаженні фахівця з інтерв'ювання і проводиться за формулою:

$$n_{min} = m \times t, \quad (17)$$

де n_{min} – мінімальний розмір територіальної одиниці (кількість домогосподарств або осіб);

m – навантаження фахівця з інтерв'ювання;

t – кількість років перебування територіальної одиниці у вибірці.

У випадку часткової ротації одиниць вибірки мінімальний розмір територіальної одиниці розраховується з урахуванням установленної схеми ротації. За умови використання у вибірковому обстеженні часткової ротації одиниць вибірки необхідно виділяти два періоди: впровадження схеми ротації (зазвичай перший рік) та підтримка схеми ротації (наступні роки). Це пояснюється тим, що у період упровадження схеми ротації необхідно відібрати більше одиниць вибірки, ніж у наступні періоди.

Мінімальний розмір територіальної одиниці з урахуванням схеми ротації розраховується за формулою:

$$n_{min} = n_1 + n_t, \quad (18)$$

де n_1 – кількість одиниць вибірки (домогосподарств або осіб), необхідних у перший рік запровадження схеми ротації, який розраховується за формулою:

$$n_1 = (q_1 \times z), \quad (19)$$

де q_1 – кількість ротаційних груп (підгруп) у перший рік запровадження схеми ротації;

z – розмір ротаційних груп (підгруп) (домогосподарств або осіб).

Кількість одиниць вибірки (домогосподарств або осіб), необхідних у наступні роки функціонування схеми ротації (n_t), розраховується за формулою:

$$n_t = (q \times z) \times (t - 1), \quad (20)$$

де q – кількість ротаційних груп (підгруп) у наступні роки функціонування схеми ротації протягом певного визначеного циклу (наприклад п'ятирічного).

Для забезпечення надійності функціонування територіальної вибірки протягом визначеної кількості років доцільно збільшити мінімальний розмір територіальної одиниці шляхом додавання резервного року. При повній щорічній ротації одиниць спостереження використовується формула:

$$n_{min} = m \times (t + 1), \quad (21)$$

а при розрахунках мінімального розміру територіальної одиниці з урахуванням схеми ротації – наступна формула:

$$n_{min} = (q_1 \times z) + (q \times z) \times t. \quad (22)$$

Наведена процедура унеможливить негативний вплив змін, які можуть відбутися протягом терміну експлуатації територіальної вибірки, а також дозволить забезпечити більш тривалий термін її експлуатації у випадку можливої затримки введення нової територіальної вибірки.

Приклад розрахунку мінімального розміру територіальної одиниці для проведення на окремих мережах респондентів двох вибірових обстежень населення визначений за двома підходами (у залежності від схеми ротації одиниць вибірки: повної щорічної ротації або часткової), з урахуванням резервного року, наведено в додатку 7.

Якщо у вимогах до формування вибірки був визначений очікуваний рівень участі (k_{rr}), мінімальний розмір територіальної одиниці коригується за формулою:

$$n'_{min} = \frac{n_{min}}{k_{rr}}; \quad (23)$$

3) за умови, що обстеження проводиться на постійній основі і план (дизайн) вибірки не зазнає суттєвих змін, розрахований мінімальний розмір територіальної одиниці доцільно скоригувати також на частку домогосподарств, непридатних до опитування.

Коефіцієнт для коригування на частку домогосподарств, непридатних до опитування, розраховується як сума середніх значень часток домогосподарств, непридатних для опитування, визначених на основі зведеної інформації щодо кількості зачинених, нежитлових приміщень та приміщень, у яких немає осіб, які проживають постійно, за результатами актуалізації списків адрес домогосподарств у попередньому циклі проведення обстежень;

4) у випадку проведення в одній територіальній одиниці декількох вибірових обстежень населення мінімальний розмір територіальної одиниці визначається сумою мінімальних розмірів, розрахованих для кожного з цих обстежень.

Територіальні одиниці, відібрані на першому ступені відбору, як правило, мають достатній розмір для проведення вибірового обстеження населення протягом кількох років. Проблема визначення мінімального розміру територіальної одиниці більш актуальна для локальних територій, які відбираються на другому ступені відбору (ВТОВ);

5) відбір несамопрезентативних територіальних одиниць.

У випадках необхідності відбору територіальних одиниць, розмір яких менше порога самопрезентативності, такі територіальні одиниці визначаються

як несамореєзентативні й уключаються до вибірки з імовірністю, пропорційною їх розміру.

Несамореєзентативні територіальні одиниці відбираються з дотриманням такої вимоги: одна така одиниця реєзентує групу територіальних одиниць, сума розмірів яких дорівнює порогу самореєзентативності.

Процедура відбору несамореєзентативних територіальних одиниць складається з таких основних етапів:

- упорядкування несамореєзентативних територіальних одиниць;
- визначення кількості територіальних одиниць, які необхідно відібрати;
- визначення кроку відбору та першої територіальної одиниці для відбору;
- відбір наступних територіальних одиниць.

Нижче наведено зміст зазначених етапів.

Упорядкування несамореєзентативних територіальних одиниць.

У межах кожної страти регіону складається впорядкований певним чином перелік несамореєзентативних територіальних одиниць. Нижче наведені можливі принципи впорядкування:

- за значенням певної ознаки (характеристики) територіальної одиниці;
- за географічним принципом (принцип "географічного серпантину");
- за абеткою відповідно до назви територіальної одиниці;
- за іншим принципом.

Останні два варіанти можуть бути застосовані при формуванні вибірки методом простого випадкового відбору.

При формуванні територіальної вибірки для проведення вибірових обстежень населення найчастіше застосовується метод систематичного відбору з імовірністю, пропорційною розміру, з попередньою стратифікацією генеральної сукупності, що вимагає впорядкування територіальних одиниць за певним принципом.

Упорядкування за ознакою (наприклад, за кількістю населення) передбачає наявність помітних розбіжностей явищ, що досліджуються, в одиницях різного розміру. Метою цього типу впорядкування є включення до вибірки як одиниць великого, так і одиниць малого розміру. Разом з тим упорядкування за принципом "географічного серпантину" (див. рис. 1) задовольняє вимозі рівномірного з географічної точки зору розміщення по території одиниць вибірки.

Беручи до уваги вищенаведене, при формуванні вибірових сукупностей для проведення вибірових обстежень населення для етапу відбору несамореєзентативних ПТОВ у межах кожної страти слід скласти перелік територіальних одиниць, ранжований у порядку зменшення кількості населення (або в порядку збільшення кількості населення). Райони слід упорядковувати не за ознакою (наприклад, за кількістю населення), а за принципом "географічного серпантину". Такий підхід забезпечить краще охоплення території регіону (області), а також дозволить урахувати природно-кліматичні особливості різних частин регіону, що важливо, наприклад, для вивчення виробництва сільськогосподарської продукції в домогосподарствах.

Застосування принципу "географічного серпантину" обов'язково передбачає використання картографічного матеріалу. За відсутністю карт можуть

використовуватися карти-схеми, але на них повинно витримуватися географічне розміщення об'єктів відносно один одного. Важливою умовою є необхідність однакової географічної орієнтації всіх карт у межах області та країни. Ураховуючи те, що всі карти, які виробляються спеціалізованими підприємствами, мають орієнтацію "північ-південь", аналогічну орієнтацію слід використовувати при проведенні серпантину. Серпантин слід проводити з лівого верхнього кута на схід, після досягнення краю карти необхідно робити поворот, потім вести серпантин зі сходу на захід паралельно попередній лінії. Обов'язковою вимогою при проведенні серпантину є відсутність "порожніх проходів", тобто при одному проході лінія має пройти щонайменше через одну територіальну одиницю. Територіальні одиниці за порядком проходження лінії включаються до списку.

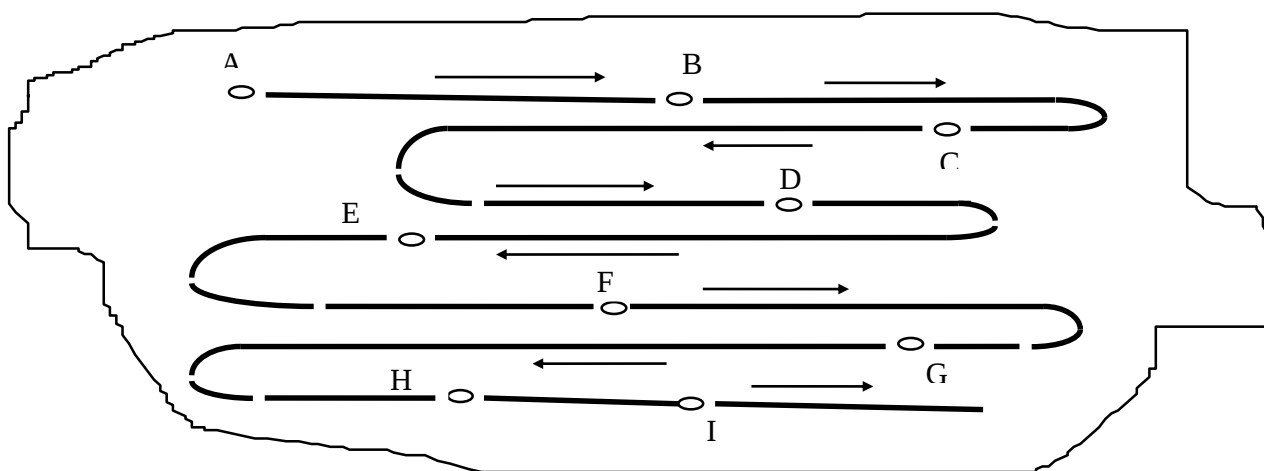


Рис. 1. Впорядковування територіальних одиниць вибірки за принципом "географічного серпантину"

Приклад упорядкування територіальних одиниць за розміром та за принципом "географічного серпантину" наведено у додатку 8.

3. Визначення кількості територіальних одиниць, які необхідно відібрати.

Кількість несаморепрезентативних ПТОВ, які необхідно відібрати у кожній страті, визначається діленням загальної кількості населення всіх несаморепрезентативних територіальних одиниць кожної страти на поріг саморепрезентативності, який визначається за формулою (15) або за формулою (13). Результатом практично в усіх випадках є дробове число, яке необхідно округлити до цілого числа. Кількість несаморепрезентативних територіальних одиниць, яку необхідно відібрати, розраховується за формулою:

$$d = N_{sstr} / n_P^{HH} = N_{sstr}^{pers} / n_P^{pers}, \quad (24)$$

де d – кількість одиниць для відбору;

N_{sstr} – кількість домогосподарств у конкретній страті;

N_{sstr}^{pers} – кількість населення у конкретній страті.

Після округлення кількості несаморепрезентативних територіальних одиниць до цілого значення визначається величина інтервалу (крок) відбору.

4. Визначення кроку відбору та першої одиниці для відбору.

Крок відбору – це результат ділення загального значення ознаки, визначеної як основа для ранжування (наприклад, кількості населення) у несаморепрезентативних територіальних одиницях області (району), на число одиниць, які необхідно відібрати.

Якщо несаморепрезентативні територіальні одиниці відранжовані за кількістю населення, крок відбору h розраховується за формулою:

$$h = N_{sstr}^{pers} / d; \quad (25)$$

5. Відбір першої територіальної одиниці вибірки.

Перед початком відбору до переліку несаморепрезентативних територіальних одиниць вносяться дані щодо кумулятивних (накопичених) значень ознаки, визначеної як основа для ранжування (наприклад, щодо кумулятивної кількості населення відповідних територіальних одиниць).

Початок відбору (перша відібрана несаморепрезентативна територіальна одиниця ($h^{(1)}$)) визначається шляхом множення кроку відбору на випадкове число:

$$h^{(1)} = h \times rnd, \quad (26)$$

де rnd – випадкове число, визначене для страти (випадкове число, яке набуває значення від 0 до 1 і може бути визначене за допомогою генератора випадкових чисел).

Отриманий результат розрахунку порівнюється з накопиченими значеннями ознаки (наприклад, кількостями населення або розмірів площ землі у переліку територіальних одиниць). Територіальна одиниця, накопиченому значенню ознаки якої відповідає результат, є першою одиницею вибірки.

6. Відбір наступних територіальних одиниць вибірки.

Усі наступні територіальні одиниці визначаються шляхом додавання кроку відбору до накопиченого значення ознаки, розрахованого для відбору попередньої територіальної одиниці:

$$h^{(z)} = h^{(z-1)} + h, \quad z = 1, 2, \dots, d, \quad (27)$$

де d – кількість несаморепрезентативних територіальних одиниць, яку необхідно відібрати;

$h^{(z)}$ – накопичене значення ознаки для відбору територіальної одиниці z .

Приклад розрахунку кроку відбору, визначення першої та наступних одиниць для відбору наведено у додатку 8.

Результати відбору з використанням двох принципів упорядкування територіальних одиниць свідчать, що територіальні одиниці, розмір яких

близький до кроку відбору, мають високу ймовірність бути відібраними за кожним з цих принципів упорядкування.

7. Розрахунок імовірностей відбору ПТОВ.

Після відбору ПТОВ розраховується ймовірність їх відбору.

У зв'язку з тим, що при формуванні територіальної вибірки для проведення вибірових обстежень населення найчастіше застосовується метод відбору з імовірностями, пропорційними розміру, нижче наведено методологічні основи до розрахунків імовірностей відбору на прикладі саме цього методу для випадку, якщо характеристикою розміру територіальної одиниці є кількість домогосподарств. У всіх територіальних одиницях проживають в цілому N домогосподарств, з яких необхідно відібрати n домогосподарств з дотриманням однакової ймовірності відбору ($f = n_{\text{факт}} / N$) для кожного домогосподарства. Це один з основних принципів, який покладено в основу методу відбору *PPS*.

Наприклад, усі домогосподарства проживають у D територіальних одиницях, з яких треба відібрати d одиниць. Відповідно кількість домогосподарств, які мають бути відібрані в кожній територіальній одиниці вибірки, дорівнює в середньому $b = n / d$.

Якщо M_α є розміром територіальної одиниці α (у цьому випадку – кількістю домогосподарств у територіальній одиниці) і $M = \sum M_\alpha$ є загальним розміром усіх територіальних одиниць D у генеральній сукупності, то ймовірність відбору територіальної одиниці α , пропорційна її розміру, становитиме [14]:

$$P_\alpha = \left(d \times \frac{M_\alpha}{\sum_\alpha M_\alpha} \right), \quad (28)$$

де P_α – імовірність відбору ПТОВ.

Наявність коефіцієнта d у формулі (28) означає, що територіальні одиниці з розміром M_α , рівним чи більшим певного порогового значення, рівного $\sum M_\alpha / d$, мають імовірність відбору, рівну чи більшу одиниці. Відповідно ці територіальні одиниці відбираються з 100% визначеністю, тобто з імовірністю, що дорівнює одиниці.

В окремих випадках на ймовірність відбору мають вплив специфічні вимоги до формування територіальної вибірки. Наприклад, однією з вимог до формування територіальної вибірки для проведення вибірових обстежень населення може бути проведення ОСГД у всіх адміністративних районах сільської місцевості України. У цьому випадку останні включаються до вибірки з імовірністю, яка дорівнює одиниці.

2. Відбір вторинних територіальних одиниць вибірки

1. Формування ВТОВ.

Для формування ВТОВ необхідно мати картографічний матеріал (карти, карти-схеми, плани), який повинен містити географічне відображення меж утворень локальних територіальних одиниць, на основі яких формуються ВТОВ.

Розмір кожної ВТОВ повинен бути не меншим розрахованого мінімального розміру територіальної одиниці відбору n_{min} . Якщо розмір локальної території більший ніж n_{min} , то вона визначається як ВТОВ; якщо розмір менший ніж n_{min} , то необхідно здійснити об'єднання локальних територій, які географічно знаходяться поруч, поки не буде забезпечено мінімально необхідний розмір (кількість населення (домогосподарств)).

2. Упорядкування ВТОВ.

Упорядкування ВТОВ при проведенні відбору доцільно здійснювати за принципом "географічного серпантину". Застосування цього принципу дозволяє враховувати особливості одиниць спостереження відповідно до їх географічного розміщення. Наприклад, у містах існують райони, які відповідно до специфіки заселення та подальшого формування в побуті мають назви "робочі", "студентські", "привілейовані" тощо. У сільській місцевості, де як ВТОВ виступають сільські населені пункти, а особливості господарювання та розташування господарств населення залежать від природно-кліматичних умов, наявності транспортної інфраструктури тощо, принцип "географічного серпантину" дозволяє краще, ніж упорядкування територіальних одиниць відповідно до кількості населення, врахувати особливості, притаманні умовам життя сільського населення.

3. Відбір ВТОВ.

Відбір ВТОВ здійснюється з використанням аналогічних процедур, які застосовуються при відборі ПТОВ (див. додаток 8).

Розрахунок імовірностей відбору територіальних одиниць другого ступеня.

Імовірність відбору ВТОВ (P_β) розраховується за формулою:

$$P_\beta = \left(b \times \frac{M_\beta}{\sum_\beta M_\beta} \right), \quad (29)$$

де b – кількість ВТОВ, які необхідно відібрати;

M_β – кількість домогосподарств у ВТОВ β ;

$\sum_\beta M_\beta$ – загальна кількість домогосподарств у всіх ВТОВ.

3. Відбір домогосподарств

1. Упорядкування домогосподарств.

Упорядкування домогосподарств здійснюється у межах відібраних ВТОВ. Вибір способу впорядкування залежить від наявної інформаційної бази і має особливості у міській та в сільській місцевостях. Метод упорядкування повинен бути однаковим у всіх ВТОВ у межах певної страти.

У міській місцевості, незалежно від джерела інформації щодо адрес домогосподарств (матеріали перепису населення, дані виборчих діляниць, дані поштових відділень тощо), упорядкування здійснюється у такому порядку:

спочатку впорядковуються за абеткою назви вулиць, потім – у порядку зростання номерів будинків, потім – у порядку зростання номерів квартир.

У сільській місцевості найбільш доцільним для використання джерелом інформації щодо домогосподарств є форма погосподарського обліку № 1 "Облікова картка об'єкта погосподарського обліку". Для зменшення обсягу робіт із формування вибірки списки адрес домогосподарств доцільно складати у порядку зростання номерів об'єктів погосподарського обліку, зазначених у формі погосподарського обліку № 1 "Облікова картка об'єкта погосподарського обліку".

Для формування вибірових сукупностей можуть бути використані й інші джерела інформації щодо адрес домогосподарств. За умови наявності декількох інформаційних джерел доцільно серед них визначити для використання більш повні за обсягом, а також більш актуальні за часом створення.

Перед початком відбору домогосподарств обов'язково здійснюється актуалізація складених списків адрес, яка унеможливує або мінімізує негативний вплив можливих змін, які відбуваються як у межах ВТОВ (наприклад, знесення будинків, стихійне лихо тощо), так і безпосередньо серед домогосподарств.

Домогосподарствам у впорядкованих списках по кожній ВТОВ зазначаються порядкові номери, починаючи з одиниці.

2. Здійснення відбору одиниць спостереження – осіб/домогосподарств. Відбір одиниць спостереження проводиться способом систематичного відбору з упорядкованих їх переліків, складених у рамках відібраних для обстеження територіальних одиниць. Наприклад, відбір домогосподарств здійснюється з упорядкованого за певними ознаками (назва вулиці, номер будинку, номер квартири і т.п.) переліку адрес домогосподарств у межах відібраних для обстеження ВТОВ. Приклад проведення процедури систематичного відбору наведено при розгляді відбору несамопрезентативних територіальних одиниць (міст) (додаток 8). Ця процедура може бути використана і при відборі осіб/домогосподарств з адміністративних реєстрів.

У зв'язку із застосуванням генератора випадкових чисел результати розрахунків кроку відбору, першого та наступних номерів відібраних адрес домогосподарств становлять дробові числа, які слід округлити за правилами арифметики до певної кількості цифр після коми. Ця кількість цифр залежить від кількості домогосподарств, яку необхідно відібрати, і визначається умовою забезпечення всім домогосподарствам з основи вибірки однакової можливості бути відібраними. Для здійснення процедури відбору слід визначення номерів відібраних адрес домогосподарств здійснювати на основі цілої частини згаданого дробового числа незалежно від значення дробової частини (тобто без округлення). Наприклад, при округленні до двох знаків після коми, незалежно від того, становить розрахунковий номер адреси домогосподарства для відбору 74,00; 74,18; 74,50; 74,61 або 74,99, до вибірки включається адреса домогосподарства, яке у списку відповідає номеру 74. Цей підхід унеможливує при відборі одиниць вибірки вихід за межі списку адрес домогосподарств у випадку проведення округлення за правилами арифметики.

3. Розрахунок імовірностей відбору одиниць третього ступеня.

Останнім етапом формування вибіркової сукупності є розрахунок імовірностей відбору та базових ваг для кожної відібраної одиниці спостереження. Імовірність відбору характеризує шанс конкретного домогосподарства бути відібраним.

Імовірність відбору домогосподарств у ВТОВ розраховується за формулою:

$$P_c = \left(\frac{c_\beta}{M_\beta} \right), \quad (30)$$

де c_β – кількість домогосподарств, які необхідно відібрати у ВТОВ β .

Розрахунок загальної ймовірності відбору для методу триступеневого відбору здійснюється за такою формулою:

$$\left(d \times \frac{M_\alpha}{\sum_\alpha M_\alpha} \right) \times \left(b \times \frac{M_\beta}{\sum_\beta M_\beta} \right) \times \left(\frac{c_\beta}{M_\beta} \right) = n/N = f, \quad (31)$$

де f – частка відбору [14].

Якщо на першому та другому ступенях відбору застосовуються однакові одиниці вимірювання розміру, в ідеальному випадку повинна виконуватися умова $M_\alpha = \sum_\beta M_\beta$, однак ця умова не є необхідною.

Задовільною є умова $M_\alpha \approx \sum_\beta M_\beta$.

Загальна теоретична ймовірність для триступеневого відбору певного i -го домогосподарства (P_i) з вибіркової сукупності розраховується за формулою:

$$P_i = P_a \times P_\beta \times P_c \quad (32)$$

Якщо у вибіркового обстеженні застосовується двоступеневий відбір (наприклад для ОСГД), формула для загальної теоретичної ймовірності відбору для певного i -го домогосподарства (P_i) набуває вигляду:

$$P_i = P_a \times P_c \quad (33)$$

Розрахунки за формулами (30) – (33) можуть здійснюватися як для всієї вибірки, так і в межах кожної страти окремо.

Х. Актуалізація основи вибірки

1. Головною метою реалізації процедури актуалізації основи вибірки є підвищення її якості шляхом унесення змін, які відбулися з моменту її формування або попередньої актуалізації.

2. Діюча в органах державної статистики система організації вибіркового обстеження населення передбачає формування територіальної вибірки на певний термін (як правило, на п'ять років плюс один резервний рік). Відповідно перший та другий етапи формування вибірки також здійснюються з такою ж періодичністю. У зв'язку з цим актуалізацію основ територіальної вибірки першого та другого ступенів відбору протягом зазначених п'яти років проводити недоцільно.

3. За умов упровадження в систему організації вибіркового обстеження населення обстеження EU-SILC, у якому домогосподарства мають перебувати у вибірці певну кількість років залежно від обраної схеми ротації, може бути доцільним обрати інші терміни формування територіальної вибірки. Наприклад, для поширеної схеми ротації, за якої щорічно замінюється близько 25% домогосподарств у вибірці, найбільш раціональний термін формування територіальної вибірки становить 4 роки. Але при цьому кожного року необхідним є оновлення певної частини територіальної вибірки. Це стосується насамперед ВТОВ та невеликих за розміром ПТОВ, які мають бути замінені після їх обстеження протягом чотирьох років.

4. У випадках знесення житлових будинків або їх руйнування через стихійне лихо на території обстежуваної ВТОВ може виникнути необхідність її заміни. Разом з тим для таких випадків також недоцільно робити актуалізацію, оскільки як для етапів відбору нових міських населених пунктів, так і для етапу відбору ВТОВ слід користуватися діючою основою територіальної вибірки. Це пов'язано з необхідністю застосування для розрахунків імовірностей відбору (формули (28), (29)) єдиної інформаційної бази для всіх адміністративно-територіальних одиниць.

5. Разом з тим, у зв'язку зі змінами, які постійно відбуваються на відібраних територіях і призводять до певної невідповідності основи вибірки її реальному стану, актуалізацію списків адрес домогосподарств необхідно здійснювати перед кожним формуванням вибіркової сукупності. Актуалізація списків адрес домогосподарств є однією з основних засад, які забезпечують якісне функціонування вибіркової сукупності домогосподарств протягом усього строку дії територіальної вибірки.

6. Метою процедури актуалізації списків адрес домогосподарств, які є основою для формування нової вибіркової сукупності домогосподарств, є підвищення якості списків шляхом унесення до них змін, які відбулися з моменту складання списків або попередньої їх актуалізації.

7. Інформаційною базою для проведення актуалізації списків адрес домогосподарств є:

списки адрес домогосподарств, які були складені для попереднього відбору;

дані форми № 1 "Облікова картка об'єкта погосподарського обліку", яка ведеться у сільських населених пунктах;

дані квартирних відділів підприємств, організацій, установ (по відомих будинках); правлінь кооперативів (по кооперативних будинках); дані по окремих домовласниках та дані об'єднань співвласників багатоквартирних будинків; центрів надання адміністративних послуг, підприємств районних електромереж та управлінь газового господарства тощо;

відомості фахівців з інтерв'ювання щодо змін, які відбулися на територіях, на яких вони проводять вибіркове обстеження населення.

8. Актуалізацію здійснюють працівники територіальних органів Держстату, до функціональних повноважень яких належать питання організації вибіркових обстежень населення. До виконання польових етапів робіт з актуалізації списків залучаються фахівці з інтерв'ювання.

9. За результатами актуалізації списків адрес домогосподарств на регіональному рівні формується зведена інформація про стан списків адрес домогосподарств у розрізі ПТОВ/ВТОВ регіону, копія якої надсилається на центральний рівень.

10. Інформація, зведена за підсумками стану списків адрес домогосподарств у попередньому циклі проведення обстежень, використовується для розрахунку частки адрес домогосподарств, непридатних до опитування, для коригування мінімального розміру територіальної одиниці (додаток 7).

11. Інформація, отримана за результатами актуалізації списків адрес домогосподарств, також застосовується для щорічного аналізу якості побудови основи вибірки.

Контроль якості формування та актуалізації основ вибірки є обов'язковим етапом процесу формування вибіркових сукупностей респондентів.

Порядок проведення цього контролю визначається методикою формування вибіркових сукупностей для проведення певного циклу вибіркових обстежень населення.

Директор департаменту
обстежень домогосподарств

Інна ОСИПОВА

Додаток 1
до Методологічних положень
(пункт 1 розділу III)

**Макет документа "Вимоги до формування вибірових сукупностей
для проведення органами державної статистики
вибіркових обстежень населення (домогосподарств)"**

1. Мета побудови вибірки.
2. Генеральна сукупність.
3. Методологія побудови вибірки.
4. Одиниці спостереження.
5. Вимоги щодо представлення у вибірці територіальних одиниць відповідно до адміністративно-територіального устрою України.
6. Термін дії територіальної вибірки.
7. Тип вибірки, метод та одиниці відбору.
8. Кількість територіальних одиниць для відбору для проведення в них вибіркового обстеження одним фахівцем з інтерв'ювання у міській та сільській місцевостях.
9. Обмеження щодо представлення у вибірці окремих груп одиниць спостереження.
10. Перелік основних показників, надійність оцінок яких має бути забезпечена у вибіркового спостереженні.

		Назва обстеження	
11.	Навантаження фахівця з інтерв'ювання, домогосподарств:		
	у великих містах		
	у малих містах		
	у сільській місцевості		
12.	Надійність оцінювання основних показників, зазначених у пункті 10, коефіцієнт варіації (CV), %:		
		назва показника	назва показника
	на державному рівні		
	на регіональному рівні		
13.	Очікуваний рівень участі:		
	у великих містах		
	у малих містах		
	у сільській місцевості		
14.	Термін дії вибірки респондентів		
15.	Наявність і схема ротації одиниць спостереження		
16.	Обсяг вибірки респондентів		

17. Додаткові вимоги до формування вибірки.

Додаток 2
до Методологічних положень
(пункт 6 розділу IV)

**Схема ротації домогосподарств
в обстеженні робочої сили**

Рік	Квартал	Місяць	№ обстеження ¹	Ротаційні підгрупи			
1	2	3	4	5	6	7	8
2019	1	січень	1901	011	021	051	061
		лютий	1902	012	022	052	062
		березень	1903	013	023	053	063
	2	квітень	1904	021	031	061	071
		травень	1905	022	032	062	072
		червень	1906	023	033	063	073
	3	липень	1907	031	041	071	081
		серпень	1908	032	042	072	082
		вересень	1909	033	043	073	083
	4	жовтень	1910	041	051	081	091
		листопад	1911	042	052	082	092
		грудень	1912	043	053	083	093
2020	1	січень	2001	051	061	091	101
		лютий	2002	052	062	092	102
		березень	2003	053	063	093	103
	2	квітень	2004	061	071	101	111
		травень	2005	062	072	102	112
		червень	2006	063	073	103	113
	3	липень	2007	071	081	111	121
		серпень	2008	072	082	112	122
		вересень	2009	073	083	113	123
	4	жовтень	2010	081	091	121	131
		листопад	2011	082	092	122	132
		грудень	2012	083	093	123	133
2021	1	січень	2101	091	101	131	141
		лютий	2102	092	102	132	142
		березень	2103	093	103	133	143
	2	квітень	2104	101	111	141	151
		травень	2105	102	112	142	152
		червень	2106	103	113	143	153
	3	липень	2107	111	121	151	161
		серпень	2108	112	122	152	162
		вересень	2109	113	123	153	163
	4	жовтень	2110	121	131	161	171
		листопад	2111	122	132	162	172
		грудень	2112	123	133	163	173

¹ У графі "Номер обстеження" перші дві цифри позначають рік проведення обстеження, а дві наступні – місяць проведення обстеження.

Продовження додатка 2

1	2	3	4	5	6	7	8
2022	1	січень	2201	131	141	171	181
		лютий	2202	132	142	172	182
		березень	2203	133	143	173	183
	2	квітень	2204	141	151	181	191
		травень	2205	142	152	182	192
		червень	2206	143	153	183	193
	3	липень	2207	151	161	191	201
		серпень	2208	152	162	192	202
		вересень	2209	153	163	193	203
	4	жовтень	2210	161	171	201	211
		листопад	2211	162	172	202	212
		грудень	2212	163	173	203	213
2023	1	січень	2301	171	181	211	221
		лютий	2302	172	182	212	222
		березень	2303	173	183	213	223
	2	квітень	2304	181	191	221	231
		травень	2305	182	192	222	232
		червень	2306	183	193	223	233
	3	липень	2307	191	201	231	241
		серпень	2308	192	202	232	242
		вересень	2309	193	203	233	243
	4	жовтень	2310	201	211	241	251
		листопад	2311	202	212	242	252
		грудень	2312	203	213	243	253

Додаток 3
до Методологічних положень
(пункт 6 розділу IV)

Приклад схеми ротації вибірки в обстеженні EU-SILC

При реалізації схеми ротації з 4 ротаційних груп зі щорічною заміною однієї ротаційної групи (близько 25% домогосподарств) вибірка для будь-якого року складається з чотирьох ротаційних груп. В усталеному процесі ротації будь-яка ротаційна група залишається в обстеженні протягом чотирьох років (рис. 1 до цього додатка). При цьому для певного року T до вибірки входять одна ротаційна група з року $T-3$, одна група з року $T-2$, одна група з попереднього року $T-1$ і одна ротаційна група, що була включена до вибірки у році T .

Щороку з вибірки видаляється "найстаріша" ротаційна група (у прикладі на рис. 2.1.1 для року T видалена група номер $i+3$, що увійшла до вибірки в році $T-4$) і додається нова (у прикладі це група номер $i+7$). Між будь-якими двома суміжними роками перекриття вибірки (обсяг вибірки у панелі) складає близько 75% домогосподарств; між будь-якими роками T і $T+2$ або $T-2$ і T перекриття вибірки становить 50%; між будь-якими роками T і $T+3$ або $T-3$ і T перекриття вибірки становить близько 25% домогосподарств (див. рис. 1 до цього додатка). Для більш тривалих інтервалів між роками перекриття дорівнює нулю.

Ротаційна група	Роки								
	T-4	T-3	T-2	T-1	T	T+ 1	T+2	T+3	T+4
i	i								
i+1	i+1	i+1							
i+2	i+2	i+2	i+2						
i+3	i+3	i+3	i+3	i+3					
i+4		i+4	i+4	i+4	i+4				
i+5			i+5	i+5	i+5	i+5			
i+6				i+6	i+6	i+6	i+6		
i+7					i+7	i+7	i+7	i+7	
i+8						i+8	i+8	i+8	i+8
i+9							i+9	i+9	i+9
i+10								i+10	i+10
i+11									i+11

Рис. 1. Схема ротації вибірки (усталений процес ротації) в обстеженні EU-SILC зі щорічною заміною однієї ротаційної групи (близько 25% домогосподарств у вибірці)

Продовження додатка 3

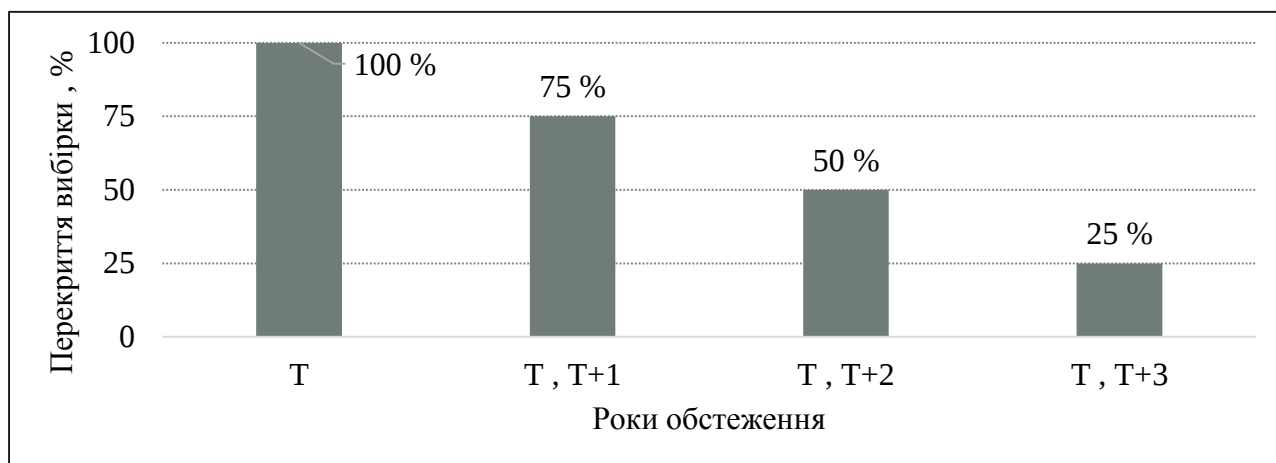


Рис. 2. Схема перебування домогосподарств у панелі вибіркової сукупності обстеження EU-SILC

Слід звернути увагу на те, що при формуванні вибірки слід урахувати вибуття домогосподарств із панелі через відмови від подальшої участі в обстеженні, переїзд, смерть однаків тощо. Для забезпечення необхідного обсягу поперечної вибірки кожного року слід формувати додаткову вибірку домогосподарств і осіб.

Одним з підходів до цього є відповідне збільшення обсягу ротаційної групи, що входить до обстеження кожного року. Обсяг фактичної вибірки такої ротаційної групи збільшується так, щоб компенсувати випадки неотримання інформації через будь-які причини у трьох ротаційних групах, що залишаються в обстеженні.

При впровадженні розглянутої схеми ротації вибірки необхідно передбачити певний період переходу до усталеного процесу ротації. Один з варіантів реалізації початкового етапу схеми ротації представлено на рис. 3 до цього додатка. На початку відбирається поперечна репрезентативна вибірка домогосподарств, яка представлена чотирма незалежними підвибірками, кожна з яких є репрезентативною щодо генеральної сукупності домогосподарств, і які є ідентичними за дизайном.

За такого підходу до впровадження схема ротації перша ротаційна група бере участь лише в одному, першому році обстеження EU-SILC. Домогосподарствам та особам, відібраним до другої ротаційної групи, пропонується взяти участь в обстеженні (у панелі) протягом двох років, до третьої ротаційної групи – протягом трьох років, а до четвертої – протягом чотирьох років. Починаючи з другого року обстеження, кожного року вводиться одна нова ротаційна група домогосподарств, яким пропонується взяти участь в обстеженні протягом чотирьох наступних років.

Ротаційна група	Початкові роки впровадження						
	1	2	3	4	5	6	...
1	1						
2	2	2					
3	3	3	3				
4	4	4	4	4			
5		5	5	5	5		
6			6	6	6	6	
7				7	7	7	7
8					8	8	8
9						9	9
10							10

Рис.3. Приклад впровадження схеми ротації вибірки до досягнення усталеного процесу ротації в обстеженні EU-SILC

Додаток 4
до Методологічних положень
(пункт 4 розділу VI)

Приклад оцінки обсягу вибірки обстеження EU-SILC в Україні

Ефективні обсяги вибірки обстеження EU-SILC в країнах ЄС представлені в табл. 1 до цього додатка.

Таблиця 1.

Мінімальний ефективний обсяг вибірки домогосподарств в обстеженні EU-SILC в країнах ЄС

Держави-члени ЄС	Домогосподарства	
	Поперечні	Поздовжні
Бельгія	4 750	3 500
Болгарія	4 500	3 500
Чеська Республіка	4 750	3 500
Данія	4 250	3 250
Німеччина	8 250	6 000
Естонія	3 500	2 750
Ірландія	3 750	2 750
Греція	4 750	3 500
Іспанія	6 500	5 000
Франція	7 250	5 500
Хорватія	4 250	3 250
Італія	7 250	5 500
Кіпр	3 250	2 500
Латвія	3 750	2 750
Литва	4 000	3 000
Люксембург	3 250	2 500
Угорщина	4 750	3 500
Мальта	3 000	2 250
Нідерланди	5 000	3 750
Австрія	4 500	3 250
Польща	6 000	4 500
Португалія	4 500	3 250
Румунія	5 250	4 000
Словенія	3 750	2 750
Словаччина	4 250	3 250
Фінляндія	4 000	3 000
Швеція	4 500	3 500
Великобританія	7 500	5 750

Доцільно розглянути приклад розрахунку обсягу вибірки для України за наведеним у розділі VI алгоритмом. Слід констатувати, що з формул (9) і (10) витікає, що обсяг вибірки не залежить від рівня показників, а цілком визначається обсягом генеральної сукупності та параметрами a і b , наведеними у табл. 1 (розділ VI).

Для показників доходів та умов життя вихідними параметрами за даними державної статистики є такі:

Кількість приватних домогосподарств, $N = 14,882$.

Ризик бідності (за критерієм ЄС), $\hat{p} = 0,10$.

Рівень соціального відчуження, $\hat{p} = 0,14$.

Стандартна похибка:

$$\widehat{SE}(\hat{p}) \leq \sqrt{\frac{\hat{p}(1-\hat{p})}{f(N)}} = \sqrt{\frac{0,10 \times (1-0,10)}{900 \times \sqrt{14,882+2600}}} = 0,0038.$$

Для забезпечення цієї умови обсяг ефективної вибірки на національному рівні має задовольняти умові:

$$n_{eff} \geq \frac{\hat{p}(1-\hat{p})}{\widehat{SE}(\hat{p})^2} = f(N) = \frac{0,10 \times (1-0,10)}{0,0038^2} = 6233 \text{ домогосподарства}$$

За таких умов граничні похибки та коефіцієнти варіації для цих показників становлять $ME(\hat{p}) = 0,008$; $CV(\hat{p}) = 3,85\%$ і $ME(\hat{p}) = 0,009$; $CV(\hat{p}) = 3,18\%$ відповідно.

На основі порівняння результатів розрахунків обсягів вибірки з даними для інших країн та з урахуванням доцільності використання схеми ротації з чотирма ротаційними групами для України слід розглядати можливість упровадження обстеження EU-SILC з ефективним обсягом поперечної вибірки $n = 6400$ домогосподарств, що приблизно відповідає обсягу вибірки, визначеному, наприклад, для Іспанії (див. табл. 1 до цього додатка).

При цьому для регіонального рівня ефективний обсяг поперечної вибірки становитиме в середньому 256 домогосподарств. Це забезпечує середню точність прямої оцінки, наприклад, показників ризику бідності для регіонів на рівні $CV(\hat{p}) = 10\%$. Вочевидь, для районів, навіть з урахуванням їх укрупнення за новим адміністративно-територіальним поділом, прямі оцінки показників будуть недостатньо надійними для їх публікації та використання.

Продовження додатка 4

Обсяг вибірки для поздовжньої складової обстеження для будь-якої пари послідовних років складається з кількості домогосподарств, які успішно пройшли інтерв'ю у першому році і в яких усі або принаймні більшість членів домогосподарств віком 16 років і старше успішно пройшли інтерв'ю в обох суміжних роках. Якщо для України оцінений ефективний обсяг вибірки для щорічної поперечної вибірки становить 6400 домогосподарств, то впровадження схеми ротації матиме вигляд, як це представлено на рис. 1 цього додатка.

Ротаційна група	Початкові роки впровадження						
	1	2	3	4	5	6	...
1	1600						
2	1600	1600					
3	1600	1600	1600				
4	1600	1600	1600	1600			
5		1600	1600	1600	1600		
6			1600	1600	1600	1600	
7				1600	1600	1600	...
8					1600	1600	...
9						1600	...
10							...

Рис. 1. Розподіл обсягу ефективної вибірки при впровадженні схеми ротації вибірки обстеження EU-SILC в Україні

Як видно з рис. 1, щороку для поперечної складової обсяг ефективної вибірки становить 6400 домогосподарств (4 ротаційні групи). При цьому кожні два послідовні роки для поздовжньої складової обсяг вибірки становить 4800 домогосподарств (три ротаційні групи). Кожні три послідовні роки теоретичний обсяг ефективної вибірки становить 3200 домогосподарств і кожні чотири послідовні роки – 1600 домогосподарств. Доцільно звернути увагу, що це ефективний обсяг вибірки.

Коригування ефективного обсягу вибірки з метою визначення необхідного досягнутого обсягу вибірки здійснюється з урахуванням дизайн-ефекту. Для обстеження EU-SILC величина дизайн-ефекту може бути оцінена за результатами ОУЖД. Так, за даними цього обстеження, у 2019 році величина *deff* на національному рівні становить в середньому 2,2 для основних показників.

Таким чином, з урахуванням зазначеної величини дизайн-ефекту обсяг досягнутої поперечної вибірки в обстеженні EU-SILC для національного рівня має становити близько:

$$n_d = 2,2 \times 6400 = 14080 \text{ домогосподарств}$$

Для визначення фактичного обсягу вибірки досягнутий обсяг вибірки має бути скоригований з урахуванням очікуваного рівня участі домогосподарств в обстеженні k_{rr} . Коригування здійснюється за формулою:

$$n_{\text{факт}} = \frac{n_d}{k_{rr}}$$

Середній рівень участі домогосподарств в обстеженні ОУЖД у 2019 році становив 70,6% ($k_{rr} = 0,706$). Якщо прийняти, що аналогічний рівень участі домогосподарств спостерігатиметься і в обстеженні EU-SILC, то кількість домогосподарств для відбору в цьому обстеженні становитиме близько 20000 домогосподарств:

$$n_{\text{факт}} = \frac{14080}{0,706} = 19946 \text{ домогосподарств}$$

Обсяги вибірки для регіонального рівня наведені у табл. 2. Слід зазначити, що ці обсяги отримані з урахуванням необхідності забезпечення рівномірного навантаження фахівців з інтерв'ювання в межах регіонів та типів місцевості, що і зумовило зміну загального обсягу вибірки з 19946 до 19998 домогосподарств. Середній обсяг вибірки для регіонів України становить 451 домогосподарство.

Виконані експертні дослідження засвідчили, що обстеження EU-SILC може бути впроваджене в Україні без суттєвої реорганізації існуючої мережі інтерв'юерів. Ураховуючи здійснені оцінки обсягу фактичної (початкової) вибірки в обстеженні EU-SILC та його проведення протягом приблизно трьох місяців (протягом другого кварталу кожного року) додаткове навантаження на інтерв'юера становитиме близько 14 домогосподарств на місяць у великих містах (м. Київ – 16 домогосподарств), 9 домогосподарств на місяць у малих містах і 4 домогосподарства у сільській місцевості.

Наведені обсяги вибірки в обстеженні EU-SILC мають розглядатися лише як приклад і мають бути уточнені при розробці методики формування вибірки обстеження EU-SILC з урахуванням вимог до цього обстеження.

Таблиця 2.

Оцінка розподілу початкової вибірки в обстеженні EU-SILC
(з урахуванням необхідності забезпечення рівномірного навантаження на фахівців з інтерв'ювання)

Регіон	Усього	Великі міста	Малі міста	Сільська місцевість
Україна	19 998	8 118	5 208	6 672
Вінницька	756	252	168	336
Волинська	654	126	252	276
Дніпропетровська	1386	882	168	336
Донецька	826	294	364	168
Житомирська	678	210	168	300
Закарпатська	506	126	140	240
Запорізька	944	420	224	300
Івано-Франківська	578	126	224	228
Київська	898	126	448	324
Кіровоградська	738	294	168	276
Луганська	850	126	532	192
Львівська	812	252	224	336
Миколаївська	824	336	224	264
Одеська	1086	546	168	372
Полтавська	770	294	140	336
Рівненська	494	126	140	228
Сумська	674	294	140	240
Тернопільська	532	168	112	252
Харківська	1270	630	280	360
Херсонська	828	252	252	324
Хмельницька	588	168	168	252
Черкаська	672	252	168	252
Чернівецька	460	168	112	180
Чернігівська	734	210	224	300
м. Київ	1440	1440	0	0

Приклад матеріалів та даних для формування документа "Вимоги до формування вибірових сукупностей для проведення органами державної статистики вибіркового обстеження EU-SILC"

1. Опис генеральної сукупності

Генеральною (базисною) сукупністю обстеження EU-SILC є всі неінституціональні (приватні) домогосподарства та особи, члени цих домогосподарств, які постійно проживають на території України протягом періоду збору даних. Особи, які проживають у колективних домогосподарствах та в інституційних закладах, до генеральної сукупності не включаються. До генеральної сукупності обстеження також можуть не включатися домогосподарства та особи, які проживають на тимчасово окупованих територіях Автономної республіки Крим, Донецької та Луганської областей, а також у населених пунктах, розташованих на територіях радіологічного забруднення, віднесених до зони відчуження (І зона) та зони безумовного (обов'язкового) відселення (ІІ зона) внаслідок катастрофи на Чорнобильській АЕС.

2. Періодичність проведення обстеження

Поперечні й поздовжні дані в обстеженні EU-SILC збираються, обробляються та публікуються на щорічній основі. Передбачається необхідність публікації поперечних даних приблизно в середині року, наступного за роком збору даних.

Польовий етап обстеження EU-SILC, як правило, реалізується наприкінці першого та протягом другого кварталу.

3. Характеристика дизайну вибірки

Дані EU-SILC, як і дані будь-якого державного вибіркового обстеження населення домогосподарств та осіб, мають ґрунтуватися на ймовірнісній вибірці. Як правило, одиниці, що беруть участь в обстеженні, мають бути тими ж одиницями, що були відібрані для цієї мети відповідно до дизайну вибірки, тобто не мають замінюватися іншими одиницями.

Упровадження обстеження EU-SILC в Україні доцільно здійснювати на основі інтегрованого дизайну вибірки. За такого дизайну поперечні та поздовжні дані надходять з одного джерела – власне з обстеження EU-SILC, тобто і поперечні і поздовжні дані отримуються на основі фактично одного вибіркового обстеження. За такого дизайну вибірки поперечні оцінки для певного року фактично не залежать від схеми ротації одиниць у вибірці. У загальному випадку щорічні поперечні оцінки отримуються на основі незалежних підвбірок. Статистична перевага такого дизайну полягає також в тому, що дані можуть накопичуватися протягом декількох років проведення обстеження для отримання кумулятивних вибірок більшого обсягу, що дозволяє проводити більш детальний

аналіз і, зокрема, забезпечує можливість більшої просторової деталізації при розрахунку регіональних та субрегіональних оцінок показників.

4. Одиниці спостереження

В обстеженні EU-SILC дані збираються за чотирма основними блоками (Methodological guidelines and description of EU-SILC):

дані щодо домогосподарств;

дані щодо складу домогосподарства та основні характеристики членів домогосподарства;

дані щодо доходів, освіти, зайнятості (основні змінні), що вимірюються на рівні осіб, але, як правило, агрегуються для побудови змінних на рівні домогосподарства;

дані, що збираються й аналізуються на рівні осіб, (детальні змінні): здоров'я, доступ до охорони здоров'я, детальна інформація про працю, історія зайнятості тощо.

Одиницями спостереження при зборі даних по перших двох із зазначених блоків є домогосподарства. При цьому дані збираються, як правило, від одного, належним чином визначеного, респондента в кожному домогосподарстві з вибірки.

Дані третього блоку мають збиратися безпосередньо на рівні осіб, охоплюючи всіх осіб в кожному домогосподарстві з вибірки. Відповідно одиницями спостереження при зборі цих даних є особи. У більшості країн ці дані збираються на основі особистих інтерв'ю за участю всіх дорослих осіб віком 16 років і старше в кожному домогосподарстві з вибірки. Опитування поєднується з детальними даними четвертого блоку, які також мають збиратися на рівні осіб.

Типи домогосподарств в обстеженні EU-SILC.

В обстеженні EU-SILC домогосподарство з вибірки циклу x вважається "відокремленим", якщо його члени з вибірки з циклу x проживають на час проведення циклу $x+1$ у більш ніж одному приватному домогосподарстві в межах території, що підлягає обстеженню. Після відокремлення одне (і тільки одне) з новоутворених домогосподарств відповідає визначенню "первісне" домогосподарство, у той час як одне або більше інших домогосподарств називаються "відокремленими" домогосподарствами.

Необхідно дотримуватися такого підходу, щоб розрізняти "первісні" і "відокремлені" домогосподарства:

якщо будь-яка особа з вибірки циклу x досі живе за тією ж адресою, що і при останньому циклі, то його/її домогосподарство визначається як "первісне" домогосподарство. Усі особи з вибірки, які переїхали, утворюють одне або декілька "відокремлених" домогосподарств;

якщо за певною адресою на час проведення циклу $x+1$ не проживає жодної особи з вибірки циклу x , то домогосподарство особи з вибірки, яка мала найменший індивідуальний номер (код) у реєстрі на час проведення останнього

циклу, вважається "первісним" домогосподарством. Якщо особи вже немає в живих або вона не проживає в приватному домогосподарстві в межах території країни, "первісним" домогосподарством буде те домогосподарство, особа з вибірки якої має найменший індивідуальний номер.

В обстеженні EU-SILC допускається використання двох видів вибірок при зборі інформації блоку 4:

початкова вибірка з "повних" домогосподарств, тобто вона охоплює всіх осіб, в кожному домогосподарстві; серед них тільки особи віком 16 років і старше на час проведення обстеження відповідають критеріям для проведення докладного особистого інтерв'ю;

випадкова вибірка осіб; при цьому тільки особи віком 16 років і старше на час проведення обстеження відповідають критеріям для проведення докладного особистого інтерв'ю.

Обидві ці схеми призначені для представлення всієї цільової сукупності осіб (а отже, і всіх приватних домогосподарств) на момент здійснення обстеження. Вони відрізняються тільки типом вибірки, що відбирається з цієї сукупності.

Особи, відібрані для збору даних за блоком 4, називаються особами з вибірки. Це всі особи або підмножина осіб з вихідної вибірки, які відстежуються протягом тривалості панелі для отримання поздовжньої вибірки обстежень. Таким чином, у принципі всі члени домогосподарств у вихідній вибірці "повних" домогосподарств є особами з вибірки.

Правила відстеження одиниць спостереження.

Основною метою поздовжньої складової обстеження EU-SILC є вимірювання змін з плином часу на рівні осіб, наприклад, перехід від навчання до роботи і від роботи до пенсії, зміни економічної діяльності і роботи, зміни у рівні доходів і бідності окремих осіб і домогосподарств. Як наслідок, у поздовжній складовій EU-SILC необхідно відстежувати осіб протягом періоду не менше чотирьох років. Поздовжні обстеження вимагають набору процедур, які визначають і вказують, хто відстежується і проходить інтерв'ю у поздовжньому обстеженні.

У кожному циклі поздовжня складова даних EU-SILC в ідеалі має бути репрезентативною щодо поточної цільової сукупності, тобто сукупності з усіх осіб, які проживають у приватних домогосподарствах у межах країни. На практиці цільова сукупність, яка фактично репрезентується поздовжньою складовою обстеження EU-SILC, відрізняється за деякими аспектами від генеральної сукупності обстеження, через спосіб побудови поздовжньої вибірки. Мета правил відстеження зводиться до відображення будь-яких змін у цільовій сукупності у вихідній вибірці й відстеженні осіб із плином часу. Вибірка для поздовжнього компонента EU-SILC у будь-який момент часу (рік), як правило, складається з наступного відстеження вихідних вибірок, відібраних раніше, а також будь-якої нової "вихідної вибірки", відібраної на момент, про який ідеться. Остання

охоплює як ротаційні групи, так і будь-які вибіркові сукупності, що можуть періодично додаватися до вибірки, щоб компенсувати зменшення панелі.

Таким чином, залежно від правил відстеження поздовжня вибірка у будь-який момент часу може не точно відображати поточну "поперечну" цільову сукупність. Тип демографічних змін, який необхідно відобразити, включає народження фізичних осіб у первісній сукупності, перехід осіб з первісної сукупності (з колективних домогосподарств, установ або переїзд за кордон) до приватних домогосподарств, що містять осіб із цієї сукупності, а також до нових приватних домогосподарств, які не містять таких фізичних осіб. За можливим винятком для доповнень до вибірки остання категорія внутрішніх іммігрантів, як правило, не поширюється на процедури відстеження панелі. Із сукупності відраховуються фізичні особи, які померли, вибули з генеральної сукупності (виїхали за кордон або перебувають за межами приватного сектору домогосподарств), або які є непридатними для обстеження з інших причин.

В ідеалі всі особи з вибірки, яких колись було відібрано, мають відстежуватися на будь-якому новому місці, куди вони переїхали. Однак, з метою зменшення витрат та з інших практичних міркувань, у EU-SILC відстежуються особи, які переїжджають тільки в межах цільової сукупності: тобто обстеженням охоплюється особа, що залишається або переїжджає в межах приватних домогосподарств на території країни обстеження. Особи з вибірки, які переїжджають до колективного домогосподарства або установи, переїжджають до територій, які не охоплюються обстеженням або переїжджають за кордон, як правило, не відстежуються. Єдиним винятком є продовження відстеження тих осіб, які переїжджають тимчасово (реальна або очікувана тривалість відсутності становить менше 6 місяців) до колективного домогосподарства або установи в межах території країни, що охоплюється обстеженням, які все ще вважаються членами відповідних домогосподарств.

Поздовжня вибірка має бути репрезентативною щодо всіх вікових груп у сукупності. Це означає, що, в принципі необхідно відстежувати осіб усіх вікових категорій. Тим не менше для цілей зменшення витрат та з інших практичних міркувань окреме відстеження може бути обмежене віком осіб. Вибір відповідного відсікання віку залежить від прийнятого дизайну вибірки EU-SILC. Мінімальні вимоги EU-SILC передбачають відстеження осіб з поздовжньої вибірки протягом чотирьох років. Для панелей такої короткої тривалості прийнятним є відстеження тільки осіб віком від 14 років і старше на момент відбору початкової вибірки для панелі.

Практичним ефектом такого обмеження є те, що діти віком до 14 років у початковій вибірці не будуть охоплюватися у поздовжній вибірці. Ці діти охоплюються тільки у випадку, якщо вони переїжджають зі свого первісного домогосподарства до нового домогосподарства, яке не містить жодної особи віком 14 років і старше. Оскільки до домогосподарств у поздовжній вибірці

включаються всі приватні домогосподарства, що містять щонайменше одну особу з вибірки, якщо відстеження обмежується особами з вибірки, старшими за певний вік, отримана вибірка не буде ідеальною через виключення домогосподарств, які містять тільки особу (осіб) з вибірки віком меншим, ніж це вікове обмеження (старші особи з вибірки будуть відсутні).

Крім того, для точного відображення демографічних змін у сукупності також необхідно передбачити включення до вибірки новонароджених дітей. Це досягається шляхом урахування народжених дітей для жінок з вибірки як осіб з вибірки і подальшим їх відстеженням за допомогою звичайних процедур. Для коротких панелей тривалістю чотири роки новонароджені діти не відстежуються.

Зміст таких обмежень щодо відстеження дітей зводиться до того, що їх поздовжню (стійку) бідність не можна оцінити точно.

В обстеженні EU-SILC (тільки для поздовжньої складової) можуть також відстежуватись колишні члени домогосподарства, щодо яких можуть збиратися деякі елементарні відомості про статус діяльності і час, проведений у домогосподарстві протягом звітного періоду обліку доходів.

5. Заміни одиниць спостереження

Для поперечної складової обстеження, а також для повздовжньої складової для домогосподарств, які обстежуються перший рік, контрольовані заміни одиниць спостереження дозволяються тільки в тих випадках, коли рівень участі в обстеженні становить менше 60% і при цьому виникла одна з таких ситуацій:

за адресою не вдалося встановити контакт, оскільки не можна знайти її місцезнаходження або вона не є доступною;

за адресою вдалося встановити контакт, але інтерв'ю не завершено через відмову домогосподарства співпрацювати, тимчасову відсутність усього домогосподарства або нездатність домогосподарства надати відповідь.

Домогосподарство, яке відмовляється проходити інтерв'ю, має бути видалене з вибірки. Будь-які особи з вибірки з такого домогосподарства автоматично видаляються з подальшого відстеження. Для панелей тривалістю чотири роки домогосподарство, яке не було включене до обстеження протягом двох років поспіль або з яким не вдалося встановити контакт у перший рік панелі, може бути видалене разом з усіма особами з вибірки в ньому. Домогосподарства, які не включені до обстеження протягом одного року через неможливість встановлення адреси, адресу, за якою ніхто не проживає, втрату даних (інформація про те, що трапилося з домогосподарством, відсутня), можуть бути видалені з обстеження. Як загальну рекомендацію можна використовувати таку: видаляються тільки ті домогосподарства, з якими не вдалося провести інтерв'ю три рази поспіль.

Заміни одиниць спостереження не допускаються для другого і наступних циклів панелі у поздовжній складовій.

При реалізації замін необхідно дотримуватися спеціальних процедур для забезпечення того, що процес замін контролюється максимально можливою мірою. Ці процедури передбачають використання плану, який гарантує, що заміни здійснюються так, щоб одиниці, включені до вибірки на заміну, максимально точно відповідали замінені одиницям за основними характеристиками. Заміни допускаються тільки після того, як було докладено спеціальних зусиль для того, щоб переконатися, що відмова відбулася насправді, а також належним чином були здійснені заплановані заходи для отримання даних від відібраних одиниць. Домогосподарство не може замінюватися іншим домогосподарством, яке проживає за тією ж адресою і не входить до вибірки.

Приклад розрахунку порога саморепрезентативності

Вхідна інформація для розрахунку порога саморепрезентативності для малих міст Вінницької області:

обсяг генеральної сукупності $N = 160\,396$ домогосподарств;
обсяг вибіркової сукупності $n_{\text{факт}} = 102$ домогосподарства;
навантаження фахівця з інтерв'ювання $m = 17$ домогосподарств;
середній розмір домогосподарства $s = 2,47$ особи.

Поріг саморепрезентативності (n_p^{pers}):

$$n_p^{\text{pers}} = \frac{N}{n} \times m \times S = \frac{160396}{102} \times 17 \times 2,47 \approx 66\,000 \text{ осіб.}$$

Таким чином, територіальні одиниці (малі міста) з населенням $n \geq 66$ тис. осіб визначаються як саморепрезентативні й відповідно включаються у вибірку з імовірністю, рівною одиниці.

Додаток 7
до Методологічних положень
(пункт 1, 2 розділу IX)

Приклад розрахунку мінімального розміру ВТОВ для проведення на окремих мережах респондентів двох вибіркового обстежень населення (домогосподарств)

Вхідна інформація для розрахунку мінімального розміру ВТОВ:
навантаження фахівця з інтерв'ювання $m = 26$ домогосподарств для обстеження;

кількість років перебування територіальної одиниці у вибірці $t = 5$ років;
коригувальний коефіцієнт (частка домогосподарств, непридатних до опитування) $k_{np} = 0,12$;

кількість ротаційних груп (підгруп) у перший рік запровадження схеми ротації $q_1 = 27$;

кількість ротаційних груп (підгруп) у наступні роки функціонування схеми ротації $q = 12$;

розмір ротаційних груп/підгруп (домогосподарств або осіб (залежно від інформаційної бази, яка зазначається у Вимогах) $z = 3$

середній розмір домогосподарства у міській місцевості $s = 2,58$ особи;

Згідно з Вимогами навантаження фахівця з інтерв'ювання має розподілятися між двома ВТОВ, тому для розрахунку мінімального розміру однієї ВТОВ (n_{min}) вводимо додатковий коригувальний коефіцієнт: $k_{part} = 0,5$.

Коригувальний коефіцієнт щодо частки домогосподарств, непридатних до опитування (k_{np}) для розрахунку мінімального розміру ВТОВ, визначається на основі даних про кількість зачинених або нежитлових приміщень, а також приміщень, у яких немає осіб, які проживають постійно, на основі інформації, зведеної за підсумками стану списків адрес домогосподарств у попередньому циклі проведення обстежень.

Мінімальний розмір ВТОВ для проведення обстеження з повною щорічною ротацією розраховуємо за формулою:

$$n_{min 1} = m \times (t + 1) \times k_{part} / (1 - k_{rr}).$$

Згідно з Вимогами термін дії територіальної вибірки становить 5 років, а з урахуванням резервного року проведення обстеження – 6 років; максимальне навантаження фахівця з інтерв'ювання становить 26 домогосподарств, коригувальні коефіцієнти становлять 0,5 (частка домогосподарств у одній ВТОВ при розподілі між двома ВТОВ) та 0,12 (частка домогосподарств, непридатних до опитування). Таким чином:

$$n_{min 1} = (26 \times 6) \times 0,5 / (1 - 0,12) \approx 89 \text{ (д/г)}.$$

Мінімальний розмір ВТОВ для проведення обстеження з частковою щорічною ротацією розраховуємо за формулою:

$$n_{min 2} = (q_1 \times z) + (q \times z) \times t / (1 - k_{rr}).$$

Продовження додатка 7

Відповідно до Вимог термін дії територіальної вибірки становить 5 років, а з урахуванням резервного року проведення обстеження – 6 років; максимально можлива кількість домогосподарств у ротаційній групі становить 3 домогосподарства; для проведення обстеження в перший рік для запровадження схеми ротації відбирається 27 ротаційних підгруп, для проведення обстеження у інші роки щорічно відбирається 12 ротаційних підгруп; коригуючий коефіцієнт становить 0,12 (частка домогосподарств, непридатних до опитування). Таким чином, мінімальний розмір ВТОВ становить:

$$n_{min\ 2} = ((27 \times 3) \times 1) + ((12 \times 3) \times 5) / (1 - 0,12) \approx 297 \text{ (д/г)}.$$

Отже, мінімальний розмір ВТОВ для проведення на окремих мережах респондентів двох вибірових обстежень населення становитиме:

$$n_{min} = (297 + 89) = 386 \text{ (домогосподарств),}$$

або:

$$n_{min} = 386 \times 2,58 = 996 \text{ (осіб)}.$$

Додаток 8
до Методологічних положень
(пункти 1, 2 розділу IX)

Приклад упорядкування територіальних одиниць за розміром та за принципом "географічного серпантину", а також реалізації процедури відбору територіальних одиниць

Вхідні дані.

У певній страті із загальною кількістю населення 183 тис. осіб знаходяться 7 територіальних одиниць:

A (кількість населення – 23 тис. осіб);

B (кількість населення – 31 тис. осіб);

C (кількість населення – 24 тис. осіб);

D (кількість населення – 34 тис. осіб);

E (кількість населення – 27 тис. осіб);

F (кількість населення – 26 тис. осіб);

G (кількість населення – 18 тис. осіб).

Територіальне розміщення цих одиниць наведено на рис. 1.

Упорядкуємо ці територіальні одиниці за двома принципами:

ранжування за зменшенням кількості населення (табл. 1);

ранжування за принципом "географічного серпантину" (табл. 2).

Таблиця 1. Приклад упорядкування територіальних одиниць за розміром (кількістю населення)

Територіальна одиниця	Кількість населення (тис. осіб)	Накопичена кількість населення (тис. осіб)	Значення кількості населення для відбору (тис. осіб)
D	34	34	
B	31	65	36,46
E	27	92	
F	26	118	97,46
C	24	142	
A	23	165	158,46
G	18	183	
Усього	183	x	

Таблиця 2. Приклад упорядкування територіальних одиниць за принципом "географічного серпантину"

Територіальна одиниця	Кількість населення (тис. осіб)	Накопичена кількість населення (тис. осіб)	Значення кількості населення для відбору (тис. осіб)
A	23	23	
B	31	54	36,46
C	24	78	
D	34	112	97,46
E	27	139	
F	26	165	158,46
G	18	183	
Всього	183	x	

Визначення кількості несаморепрезентативних територіальних одиниць, які необхідно відібрати.

Додаткові вхідні дані:

поріг саморепрезентативності $n_p \approx 66$ тис. осіб;

кількість населення у страті (у вищенаведених несаморепрезентативних територіальних одиницях) $N_{sstr}^{pers} = 183$ тис. осіб.

Кількість несаморепрезентативних територіальних одиниць, яку необхідно відібрати (d):

$$d = N_{sstr}^{pers} / n_p^{pers}$$

$$d = 183000 / 66\,000 = 2,772 \approx 3.$$

Таким чином, слід відібрати 3 територіальні одиниці із 7.

Визначення кроку відбору.

Відповідно до даних попереднього прикладу ($N_{sstr}^{pers} = 183000$ тис. осіб, $d = 3$) крок відбору становить (h):

$$h = N_{sstr}^{pers} / d = 183000 / 3 = 61 \text{ тис. осіб.}$$

Визначення першої несаморепрезентативної територіальної одиниці для відбору.

Вхідні дані:

випадкове число $rnd = 0,597704$;

крок відбору $h = 61$ тис. осіб.

Значення для відбору першої територіальної одиниці (h_1):

$$h^{(1)} = h \cdot rnd = 61000 \times 0,597704 = 36459,9 \approx 36,46 \text{ тис. осіб.}$$

Отримане значення порівнюється з накопиченими кількостями населення у переліку територіальних одиниць. Територіальна одиниця, накопиченому значенню кількості якої відповідає результат, є першою територіальною одиницею вибірки. У таблицях 1 і 2 – це територіальна одиниця В.

Відбір наступних несаморепрезентативних територіальних одиниць.

Усі наступні територіальні одиниці визначаються шляхом додавання кроку відбору до накопиченої кількості населення, розрахованої для відбору попередньої територіальної одиниці:

$$h^{(z)} = h^{(z-1)} + h.$$

Приклад відбору територіальних одиниць за даними таблиць 1 та 2:

$$h^{(1)} = 36,46;$$

$$h^{(2)} = h^{(1)} + h = 36,46 + 61 = 97,46;$$

$$h^{(3)} = h^{(2)} + h = 97,46 + 61 = 158,46.$$

Таким чином, у випадку впорядкування територіальних одиниць за розміром (ранжування за зменшенням кількості населення) будуть відібрані територіальні одиниці В, F та А (табл. 1); у випадку упорядкування за принципом "географічного серпантину" – В, D, та F (табл. 2).

Список використаних джерел

1. Закон України "Про державну статистику". [Електронний ресурс]. – Режим доступу: <http://www.ukrstat.gov.ua>
2. Закон України "Про інформацію". [Електронний ресурс]. – Режим доступу: <http://www.ukrstat.gov.ua>
3. Закон України "Про доступ до публічної інформації" від 13 січня 2011 року № 2939-VI (зі змінами та доповненнями) (URL: https://zakon.rada.gov.ua/laws/show/2939_17).
4. Закон України "Про захист персональних даних" від 01 червня 2010 року № 2297-VI (зі змінами) (URL: <https://zakon.rada.gov.ua/laws/show/2297-17>).
5. Методологічні положення щодо забезпечення статистичної конфіденційності в органах державної статистики, затверджені наказом Держстату від 15 лютого 2017 року № 41 (URL: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2017/41/41_2017.htm).
6. Глосарій до плану статистичного спостереження, затверджений наказом Держкомстату від 29 грудня 2009 року № 498 [Електронний ресурс]. – Режим доступу: <http://www.ukrstat.gov.ua>
7. 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.
8. Бокун Н. Ч., Чернышова Т. М. Методы выборочных обследований: Учебно-справочное пособие. – Мн.: НИИ статистики Республики Беларусь, 1997. – 416с.
9. Єріна А. М. Організація вибірових обстежень: Навчальний посібник. – К.: КНЕУ, 2004. – 127 с.
10. Йейтс Ф. Выборочный метод в переписях и обследованиях. – М.: Статистика, 1965. – 433 с.
11. Кокрен У. Методы выборочного исследования. - М.: Статистика, 1976. – 440с.
12. Колтон Г. Структура выборки для проведения обследования домашних хозяйств на Украине // Отчет Госкомстата Украины. 1998. – 20 с.
13. Розробка та впровадження методології формування вибіркової сукупності для обстеження бюджетів домашніх господарств, її ротації та розповсюдження результатів на генеральну сукупність / Науковий звіт НДІ статистики Держкомстату України за 1998 рік.
14. Саріогло В. Г. Проблеми статистичного зважування вибірових даних: Монографія. – К.: ІВЦ Держкомстату України, 2005. – 264 с.
15. Labour force survey in the EU, candidate and EFTA countries — Main characteristics of national surveys, 2015 [Електронний ресурс.–Режим доступу: <http://ec.europa.eu/eurostat/documents/3888793/7751652/KS-TC-16-021-EN-N.pdf/8475c2e2-c037-4ba2-9029-93db1ade41fe>].
16. Household Budget Surveys in the EU Methodology and recommendations for harmonisation – 2003 [Електронний ресурс.–Режим доступу: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-BF-03-003-__-N-EN.pdf]

17. Kalton G. Introduction to Survey Sampling. Newbury Park: SAGE Publications Inc., 1983. – 96 p.

18. Kish L. Survey sampling.- Wiley Classics Library Edition Published, 1995.- 643p.

19. Methodology of European labour force surveys:(2) Sample design and implementation Francesca Gagliardi, Vijay Verma, Giulia Ciampalini Working Paper n. 79, January 2009 [Електронний ресурс.–Режим доступу: https://www.researchgate.net/publication/228877388_Methodology_of_European_labour_force_surveys2_Sample_design_and_implementation].

20. Verma V. Workshop on Labor Force Surveys for CIS Countries and the Baltic States run by the International Labor Office, Bureau of Statistics hosted by State Statistics Committee of Ukraine (Kyiv, 14-19 September 1998). Part Three: sample design, data evaluation, field operations.

21. Regulation (EC) No 2019/1700 of 10 October Regulation (EU) establishing a common framework for European statistics relating to persons and households, based on data at individual level collected from samples, amending Regulations (EC) No 808/2004, (EC) No 452/2008 and (EC) No 1338/2008 of the European Parliament and of the Council, and repealing Regulation (EC) No 1177/2003 of the European Parliament and of the Council and Council Regulation (EC) No 577/98 [Електронний ресурс.–Режим доступу: <https://op.europa.eu/en/publication-detail/-/publication/9c53922a-eed8-11e9-a32c-01aa75ed71a1/language-en>].

22. Регламент (ЄС) № 1982/2003 від 21 жовтня 2003 року, що впроваджує Регламент (ЄС) № 1177/2003 Європейського Парламенту і Ради стосовно статистики співтовариства щодо доходів і умов життя (EU-SILC) щодо формування вибірки і правил відслідковування.

23. Звіт № 4 "Розробка проєкту методики розрахунку обсягу вибірки з урахуванням впровадження EU-SILC в національну систему вибіркових обстежень населення та підготовка рекомендацій щодо створення (реорганізації) мережі інтерв'юерів", виконаний в рамках проєкту "Розвиток системи вибіркових обстежень населення відповідно до міжнародних норм і стандартів на основі використання сучасних технологій електронного збору даних і способів поширення інформації, а також удосконалення спроможності ДССУ і методологічна підтримка впровадження обстеження домогосподарств "Статистика доходів і умов життя в Європейському Союзі (EU-SILC)" / проєкт Світового банку "Розробка статистичних методик і розбудова спроможностей в умовах збройного конфлікту в Україні" (Грант № TF0A7020). – Контракт № CQ-02. – К.: 2020. – 70 с.

24. Регламент Європейського Парламенту та Ради від 11 березня 2009 року № 223/2009 "Про європейську статистику" (URL:<https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32009R0223>).