

Державна служба статистики України

ЗАТВЕРДЖЕНО

Наказ Державної служби
статистики

6 серпня 2014 р. № 226

(зі змінами, внесеними наказом
Державної служби статистики від
19 грудня 2022 р. № 324)

**МЕТОДОЛОГІЧНІ ПОЛОЖЕННЯ
ЩОДО ЗАСТОСУВАННЯ ВИБІРКОВОГО МЕТОДУ ОБСТЕЖЕННЯ В
ДЕРЖАВНИХ СТАТИСТИЧНИХ СПОСТЕРЕЖЕННЯХ СТАТИСТИКИ
ПІДПРИЄМСТВ**

Київ – 2014

Методологічні положення щодо застосування вибіркового методу обстеження в державних статистичних спостережень статистики підприємств (далі – Методологічні положення) визначають основні принципи і методи організації та проведення таких спостережень, зокрема методів побудови плану вибірки, методів оцінювання показників за результатами спостережень та методів визначення рівня надійності оцінок показників. Також в методологічних положеннях висвітлені особливості організації застосування вибіркового методу в державних статистичних спостереженнях.

Методологічні положення підготовлені за консультаційної підтримки фахівця Інституту демографії та соціальних досліджень імені М.В. Птухи НАН України, д.е.н. Саріогло В.Г. та експерта ЄС за проектом Твіннінг "Сприяння процесам удосконалення Державної служби статистики України з метою покращення її потенціалу та продукції" компоненту «Методологія вибірових обстежень» Даліуте Каваліускіне, працівниками департаменту статистичної методології Держстату України:

- Товченко А.В. – начальник відділу математичних методів оброблення та аналізу статистичних даних
- Ткаченко О.М. – головний спеціаліст-економіст відділу математичних методів оброблення та аналізу статистичних даних

Методологічні положення схвалені Комісією з питань удосконалення методології та звітної документації Держстату України (протокол від 25.07.2014 № 7).

Держстат України:
вул. Шота Руставелі 3, Київ, 01023
<http://www.ukrstat.gov.ua>
телефон: (044) 287-24-22
адреса електронної пошти: office@ukrstat.gov.ua

Зміст

	Стор.
Передмова	4
1. Загальні положення.....	5
2. Визначення основних термінів і понять	6
3. Сукупності одиниць	9
4. Методи відбору одиниць до вибіркової сукупності	9
5. Похибки вибірки та невибіркові похибки	11
6. Обсяг вибірки	12
7. Побудова плану вибірки	13
8. Методи редагування та зважування даних.....	15
9. Рекомендовані стандартні програмні продукти для проведення державних статистичних спостережень	18
Список використаних джерел	19
Додатки.....	21

Передмова

Згідно зі Стратегією розвитку державної статистики на період до 2017 року, схваленої розпорядженням Кабінету Міністрів України від 20 березня 2013 р. № 145-р, Держстат України здійснює впровадження вибірових методів обстеження в усіх державних статистичних спостереженнях (крім спостережень у сфері статистики фінансів), респондентами яких є суб'єкти малого підприємництва.

Вибірковий метод у світі по праву вважається одним з ефективних методів вивчення економічних й соціальних явищ та процесів у суспільстві. Застосування цього методу дає можливість отримати надійні оцінки таких явищ та процесів при зменшенні витрат на проведення самих спостережень та зменшенні звітного навантаження на респондента у порівнянні з обстеженнями, що проводяться на суцільній основі.

Органи державної статистики використовували й використовують вибірковий метод, перш за все, для проведення державних статистичних спостережень, об'єктом яких є чисельні сукупності одиниць, такі, наприклад, як сукупності домогосподарств або фізичних осіб, обсяг яких обчислюється мільйонами і які є достатньо однорідними за основними характеристиками.

Наразі постало питання щодо поширення використання вибірового методу для всіх державних статистичних спостережень за діяльністю малих та мікропідприємств, кількість яких в Україні також є достатньо великою.

Задля вирішення цього питання Держстатом України розпочата підготовка уніфікованої методологічної документації із застосування вибірового методу для державних статистичних спостережень, що складають статистику підприємств.

Методологічні положення забезпечують методологічну основу для практичного впровадження вибірового методу в державних статистичних спостереженнях, у яких вивчається діяльність підприємств у різних її аспектах.

1. Загальні положення

1.1. Методологічні положення підготовлені відповідно до статті 9 Закону України "Про офіційну статистику".

В основу цих Методологічних положень покладено документи Євростату Керівництво з планування і проведення статистичних обстежень підприємств [5] та Керівництво з вибіркового обстеження [7].

1.2. Методологічні положення визначають основні принципи та методи проведення державних статистичних спостережень, а також організацію застосування вибіркового методу в державних статистичних спостереженнях з питань статистики підприємств. Порядок здійснення робіт із застосування вибіркового методу обстеження в державних статистичних спостереженнях статистики підприємств наведено у додатку 1 до цих Методологічних положень.

Вибіркові методи, визначені цими Методологічними положеннями, передбачені, перш за все, для обстеження малих та мікропідприємств, розмір яких визначається відповідно до чинного законодавства, та призначені для використання при розробленні методології та організації застосування вибіркового методу в державних статистичних спостереженнях, респондентами яких є суб'єкти малого підприємництва.

Окремі аспекти цих Методологічних положень можуть бути корисними при організації та проведенні інших державних статистичних спостережень.

1.3. Методологічні положення призначені для державних статистичних спостережень, що базуються на принципах, закладених у Концепції розвитку державної статистики підприємств України, та відповідають Критеріям, за якими державні статистичні спостереження відносяться до статистики підприємств (затверджені наказом Держстату України від 29.12.2012 № 548).

1.4. Методи, обрані для організації державних статистичних спостережень статистики підприємств із застосуванням вибіркового методу обстеження (далі - державні статистичні спостереження), мають забезпечувати належний рівень якості отриманих даних. Відповідно ці методи мають бути обґрунтованими з позиції доцільності їх використання з метою отримання максимально

об'єктивних оцінок та економічної ефективності дослідження. Державне статистичне спостереження оцінюють за основними компонентами якості даних та процесів, такими як: відповідність, своєчасність та пунктуальність, точність (надійність), доступність та зрозумілість, послідовність та зіставність, оцінка потреб та очікувань користувачів, ефективність, витрати та навантаження на респондентів, конфіденційність, прозорість та захист.

1.5. Дані щодо генеральної сукупності повинні бути об'єктивними, точними, зрозумілими, зіставними у часі.

1.6. Результати державних статистичних спостережень мають об'єктивно відображати якісні та кількісні характеристики генеральної сукупності з допустимим рівнем похибок.

1.7. Методологічні положення можуть бути використані як для проведення нових, так і для удосконалення вже існуючих державних статистичних спостережень.

1.8. Методологічні положення можуть бути використані працівниками органів державної статистики, а також можуть бути корисними для фахівців інших державних органів, наукових організацій та установ, які проводять дослідження в галузі економіки або користуються результатами таких досліджень, інших зацікавлених користувачів.

2. Визначення основних термінів і понять

У цих Методологічних положення використовуються терміни у значеннях, наданих у Глосарії до плану статистичного спостереження [1], Положенні про Реєстр статистичних одиниць та формування основ вибірки сукупностей одиниць статистичних спостережень за діяльністю підприємств [2], Рекомендаціях щодо підготовки стандартного звіту з якості державного статистичного спостереження [11], а також у таких значеннях:

базова вага – кількість одиниць генеральної сукупності, яку за планом вибірки повинна представляти відібрана у вибірку одиниця [14], тобто це характеристика, що надається кожній одиниці з вибіркової сукупності та показує кількість одиниць генеральної сукупності, яку представляє включена до вибірки одиниця;

вибіркове статистичне спостереження – науково обґрунтований вид несучільного спостереження, при якому обстежуються не всі одиниці сукупності, а лише певна їх частина, відібрана ймовірнісним методом, а одиницею спостереження є статистична одиниця;

вибірковий метод спостереження – науково обґрунтована система правил відбору одиниць і способів характеристики сукупності досліджуваних одиниць, що вивчають. Вибірковий метод спостереження дає змогу розповсюдити висновки, отримані на основі вивчення частини сукупності (вибірки), на всю сукупність (генеральну) [10];

генеральна сукупність (основа вибірки) – сукупність усіх активних статистичних одиниць із заданими характеристиками, які планується вивчати й досліджувати при проведенні статистичного спостереження у наступному році;

дизайн-ефект (deff) – міра відносної статистичної ефективності певного плану вибірки порівняно з простою випадковою вибіркою (відношення дисперсії вибіркових оцінок показника для плану вибірки, реалізованого в спостереженні, до дисперсії оцінки цього показника, отриманої за умови використання процедури простого випадкового відбору) [10];

дисперсія вибіркових оцінок – середній квадрат відхилень оцінок показника за всіма можливими вибірками певного дизайну (реплікаціями) від середнього значення оцінки за всіма вибірками [10];

довірчий інтервал – інтервал, який покриває невідомий параметр із заданою надійністю. Довірчий інтервал є інтервальною оцінкою показника і в основному використовується при невеликому обсязі вибірки [12];

зважування – спосіб обчислення статистичних показників, який полягає в тому, що в розрахунок приймаються статистичні ваги [12];

зміщення оцінки показника – різниця між очікуваною величиною оцінки та дійсним значенням показника для всієї сукупності;

імпутація – це логічна операція, яка використовується для визначення та присвоєння або заміщення значень для відсутніх, неприпустимих та суперечливих значень, які не пройшли редагування;

кластер – група одиниць спостереження, об'єднаних за певною ознакою;

коефіцієнт варіації (відносна стандартна похибка) – відношення величини стандартної похибки вибірки до оцінки показника у відсотках.

мода – найбільш типове значення випадкової величини, що найчастіше зустрічається [12];

місцева одиниця – це підприємство або його частина, що розташована в географічно визначеному місці [2];

невідповідь – випадок, коли державне статистичне спостереження не отримало всі необхідні дані від усіх визначених одиниць сукупності статистичного спостереження [11];

нетипові одиниці (аутлаери) – одиниці спостереження, для яких значення цільової ознаки перевищує рівень порогу, встановленого для певної сукупності (страсти) законодавчо або експертним шляхом, виходячи зі співвідношення цільової ознаки зі змінною стратифікації [14];

обсяг вибірки – кількість одиниць у вибірковій сукупності. Необхідний для забезпечення репрезентативності та належної якості результатів вибіркового статистичного спостереження [10];

оцінка показника – значення показника, отримане за неповними даними щодо сукупності, яка досліджується. Може бути отримана шляхом: розповсюдження даних вибіркової сукупності на генеральну; застосування моделей; визначення значень експертами тощо [10];

підприємство – це статистична одиниця, що здійснює виробництво товарів та послуг і є самостійною у прийнятті рішень щодо використання своїх

засобів, і є найбільш незалежною з огляду на фінансові та виробничі функції [2];

план (дизайн) вибірки – організаційно-логічна модель структури вибіркової сукупності та принципи її формування. Якістю плану вибірки значною мірою визначається репрезентативність вибіркової сукупності, отриманої за визначеними у плані правилами [10];

похибка вибірки – похибка, яка виникає через те, що обстежено не всю генеральну сукупність, а тільки її частину. Її величина залежить від того, які елементи генеральної сукупності потраплять до вибірки, тобто від виду та способу відбору, варіації значення ознаки, а також від обсягу вибірки [14];

редагування – це застосування перевірок для встановлення відсутніх, неприпустимих та суперечливих значень або для визначення значень, які потенційно помилкові;

реплікації – результати отримані при багаторазовому застосуванні процедур вибіркового статистичного спостереження однакового дизайну вибірки;

ряди розподілу – упорядкований розподіл одиниць сукупності на групи за групувальною ознакою;

ротація одиниць вибірки – заміна частини діючої вибіркової сукупності (або для окремих вибіркових обстежень – її повна заміна) за розробленими схемами при реалізації чергових раундів вибіркового обстеження, яке проводиться на постійній основі;

середньоквадратична похибка – величина, що характеризує надійність оцінювання показників за результатами вибіркових обстежень за умов зміщеності оцінок. Середньоквадратична похибка визначається як сума дисперсії вибіркових оцінок показника та квадрата величини зміщення оцінки, яке є різницею між очікуваною величиною оцінки (тобто середнім із вибіркових оцінок) та дійсним значенням показника для генеральної сукупності [10];

середнє арифметичне значення ознаки – сума значень ознаки групи розділена на кількість елементів у групі [14];

стандартна похибка – середньоквадратичне відхилення вибіркових оцінок показника за всіма можливими вибірками цього дизайну від дійсного значення показника для генеральної сукупності за умови незміщеності оцінок [10];

статистична одиниця – це одиниця, через яку можна отримати несуперечливі та порівняні дані щодо адміністративних одиниць, які мають різну юридичну організацію, облікову й виробничу структуру [2];

цільова сукупність – сукупність одиниць спостереження, що складає об'єкт дослідження і підлягає статистичному вивченню [12];

страта – внутрішньо однорідна за певною ознакою частина одиниць спостереження [14];

стратифікація – процес розподілення генеральної сукупності на однорідні групи (страти) [14].

3. Сукупності одиниць

3.1. При проведенні державного статистичного спостереження використовується певний набір сукупностей одиниць. До цього набору входять: цільова, генеральна та вибіркова сукупності.

Завжди хочеться мати "ідеальну" статистичну сукупність, яка б досконало відповідала потребам користувача. Така сукупність є цільовою сукупністю. Однак на практиці її важко отримати, оскільки джерело, з якого формується сукупність, може не містити повної та актуальної інформації щодо одиниць.

3.2. Для державних статистичних спостережень сукупність формується за уніфікованими критеріями з Реєстру статистичних одиниць (РСО). Сформована сукупність з РСО представляє собою генеральну сукупність одиниць, з якої формуються вибіркові сукупності.

3.3. При формуванні та використанні сукупностей слід звертати увагу на аспекти взаємозв'язку між сукупностями, переходу від однієї до іншої. В ідеальному випадку цільова сукупність має повністю відповідати генеральній сукупності. Однак, оскільки на практиці вона часто відрізняється, існують деякі помилки, зокрема надмірного та недостатнього охоплення. Ці помилки виявляються, зокрема, при проведенні спостереження, за рахунок альтернативних джерел і розрахунків.

4. Методи відбору одиниць до вибіркової сукупності

4.1. При формуванні вибірових сукупностей відбір статистичних одиниць з генеральної сукупності визначається таким чином, щоб вибіркова сукупність найбільш адекватно представляла та характеризувала генеральну сукупність.

Рівень надійності оцінок показника залежить від методів формування вибіркової сукупності та її обсягу.

4.2. Для державних статистичних спостережень використовують метод ймовірнісного відбору одиниць.

Ймовірнісний відбір – це метод відбору, оснований на об'єктивних правилах випадкового відбору, під час якого кожна одиниця генеральної сукупності має відому ненульову ймовірність бути включеною до вибірки. Основна перевага ймовірнісного методу – це можливість визначення рівня точності оцінювання показників за результатами окремого вибіркового обстеження.

4.3. Проста випадкова вибірка – вибірка, що складається з випадково відібраних одиниць із генеральної сукупності, у якій кожна одиниця має однакову ймовірність бути включеною до вибірки, або кожна можлива комбінація однакової кількості одиниць із генеральної сукупності має однакову ймовірність утворити вибірку [7]. Проста випадкова вибірка застосовується,

головним чином, до ситуації, коли немає ніякої попередньої інформації щодо одиниць генеральної сукупності. Проста випадкова вибірка служить основою для багатьох більш складних видів вибірок та для порівняння відносної ефективності розробленої вибірки за допомогою дизайн-ефекту.

4.4. Стратифікована вибірка – вибірка, в якій генеральну сукупність розбивають на страти, що формуються за певними ознаками, а потім у межах кожної зі страт реалізуються процедури ймовірнісного відбору.

Стратифікована вибірка є основним видом вибірок для державного статистичного спостереження.

Перевагою стратифікованої вибірки є те, що за умови знаходження критеріїв стратифікації, які добре корелюють з цільовими показниками, неоднорідну сукупність вдається поділити на певну кількість однорідних підсукупностей (страт), і, як результат, отримати значно вищу точність оцінок показників, ніж при простій випадковій вибірці за однакового обсягу вибірки.

4.5. Кластерна вибірка – вибірка, для якої здійснюється ймовірнісний відбір певних частин генеральної сукупності – кластерів, а потім у межах кожного кластеру реалізуються процедури ймовірнісного відбору одиниць спостереження. Тобто обирається певна кількість кластерів, які досліджуються, а інші кластери не досліджуються.

У разі кластерної вибірки кожен кластер являє собою зменшену модель генеральної сукупності. Прикладами виділення кластерів для підприємств можуть бути підприємств за такими ознаками: певні види економічної діяльності, регіони, види економічної діяльності за регіонами, середня кількість працівників тощо. Надалі за відібраними кластерами, як правило, реалізуються процедури простого випадкового відбору. При цьому загальною ймовірністю відбору певної одиниці спостереження є доданок ймовірності відбору кластеру, до якого вона належить, та ймовірності відбору цієї одиниці в межах кластеру.

4.6. Випадкова вибірка, сформована з ймовірністю, пропорційною розміру, – вибірка, в якій ймовірність включення одиниць пропорційна їх розміру або певній характеристиці розміру.

4.7. З точки зору існування однієї вибіркової сукупності протягом певного часу, зокрема у короткотермінових державних статистичних спостереженнях, найпоширенішим методом проведення вибіркового статистичного спостереження є панельний метод. Панельний метод – спосіб збору інформації, за яким протягом певного періоду в державному статистичному спостереженні періодично опитується одна і та сама сукупність одиниць. Тобто вибіркова сукупність або певна її частина формується один раз на певний період і використовується для опитування протягом всього цього періоду.

При використанні панельного методу слід брати до уваги, що вибіркова сукупність може зменшуватись за рахунок невідповідей, що може призвести до значних похибок вибірки. З метою мінімізації таких похибок обсяг вибірки при формуванні можна збільшити на певний відсоток і в подальшому протягом певного періоду до сукупності не вносити ніяких змін.

4.8. З метою кращого охоплення обстеженням генеральної сукупності та зниження звітного навантаження на респондентів для звітних одиниць із

кількістю працівників менше за встановлену норму застосовують схеми ротації. Використання схем ротації одиниць у вибірці дозволяє підвищити рівень надійності порівнянь показників у динаміці, не перевантажуючи будь-яку окрему групу одиниць занадто довгим періодом опитування, оскільки, з одного боку, забезпечується оновлення вибіркової сукупності для кожного періоду (року, кварталу, місяця) спостереження, а з іншого – зберігається незмінною певна її частина. Крім того, використання схеми ротації дає можливість більш надійно оцінювати зміни показників між суміжними періодами (роками, кварталами, місяцями).

5. Похибки вибірки та невибіркові похибки

5.1. Отримані за результатами державного статистичного спостереження оцінки показників з високим рівнем імовірності дещо відрізняються від дійсних значень цих показників для генеральної сукупності. Ці відмінності називаються похибками.

Ступінь надійності оцінки показника визначається похибками, що походять з двох основних джерел, – похибки вибірки та невибіркові похибки.

Невибіркові похибки включають:

- похибки охоплення;
- похибки вимірювання;
- похибки невідповідей;
- похибки обробки.

Невибіркові похибки впливають як на дисперсію, так і на зміщення оцінок показника.

5.2. Похибки вибірки виникають через несупільність вибіркового статистичного спостереження і характеризують дисперсію вибіркових оцінок показників.

Близькість оцінки показника до дійсного значення має два аспекти:

- вибіркові оцінки мають малу дисперсію;
- середнє значення оцінок показника близьке до дійсного значення.

Дисперсія оцінок показника, отриманих по різних вибірках (реплікаціях), називається дисперсією вибіркових оцінок або вибірковою дисперсією. Дисперсія є критерієм, що характеризує величину відхилень від дійсного значення показника для генеральної сукупності за умов відсутності зміщення.

Квадратний корінь з дисперсії вибіркових оцінок визначає величину стандартної похибки вибірки SE . Різниця між середнім значенням оцінок показника і дійсним значенням визначає величину зміщення оцінки показника.

5.3. Для визначення статистичної надійності оцінки показника у загальному випадку – з урахуванням дисперсії та зміщення – застосовують величину, яку називають середнім квадратом похибки (середньоквадратична

похибка) і позначають MSE , що дорівнює сумі дисперсії вибірових оцінок показника та квадрату зміщення оцінки показника.

Таким чином, середній квадрат похибки оцінки показника є критерієм надійності оцінки показника та характеризує величину відхилень від дійсного значення відповідної характеристики сукупності. Дисперсія вибірових оцінок показника, оцінки показника i , відповідно, середній квадрат похибки залежать від варіації оцінки показника.

Для характеристики надійності оцінок показників застосовують також величину загальної похибки оцінки – TE , що визначається як корінь квадратний з середнього квадрату похибки. Основною характеристикою надійності оцінок показників є відносна загальна похибка оцінювання показника (коефіцієнт варіації), що визначається як відношення загальної похибки до значення оцінки показника у процентах – RTE .

Для уявлення про ступінь статистичної однорідності результатів при багаторазовому застосуванні процедур вибірового статистичного спостереження однакового дизайну вибірки (реплікацій) розраховують відносну стандартну похибку (коефіцієнт варіації).

Таким чином, для визначення надійності оцінювання показників вибірових статистичних спостережень необхідно розраховувати наступні величини:

- дисперсію вибірових оцінок показника;
- стандартну похибку вибірки;
- відносну стандартну похибку або коефіцієнт варіації (CV);
- зміщення оцінки показника;
- середній квадрат похибки оцінки показника;
- відносну загальну похибку оцінки показника.

Ураховуючи, що визначення ступеня зміщеності оцінок показників за результатами вибірових обстежень є серйозною статистичною проблемою, оскільки дійсні їх значення, як правило, невідомі, найбільш поширеними характеристиками надійності є дисперсія вибірових оцінок, стандартна похибка та коефіцієнт варіації.

6. Обсяг вибірки

6.1. Обсяг вибірової сукупності повинен забезпечити надійні результати державного статистичного спостереження. При визначенні обсягу вибірки необхідно дотримуватись компромісу між надійністю, насамперед, точністю результатів державного статистичного спостереження, вартістю обстеження та навантаженням на респондента.

- Для обчислення обсягу вибірки необхідно, перш за все, визначити:
- основні показники, які необхідно оцінювати;
- ступінь варіації досліджуваної ознаки;
- точність, яку необхідно забезпечити;

очікуваний рівень невідповідей;
обсяги доступних фінансових ресурсів.

6.2. За умов достатнього фінансування, обсяг вибірки визначають з урахуванням похибок вибірки, ступеня варіації досліджуваної ознаки по одиницях сукупності, способу відбору одиниць, очікуваного рівня невідповідей.

7. Побудова плану вибірки

7.1. Побудова плану вибірки залежить від трьох ситуацій:

переведення суцільного спостереження або спостереження основного масиву на вибірковий метод обстеження;

впровадження нового державного статистичного спостереження;

удосконалення вже існуючого державного статистичного спостереження.

Для переведення суцільного спостереження або спостереження основного масиву на вибірковий метод обстеження використовується така схема побудови плану вибірки:

1. Оцінка якості даних Реєстру статистичних одиниць у частині повноти та точності необхідної інформації (класифікаційні ознаки, показники тощо). Якість даних Реєстру статистичних одиниць визначає якість формування генеральної сукупності обстеження.

2. Формування генеральної сукупності з Реєстру статистичних одиниць та аналіз часового виміру, до якого відноситься інформація щодо одиниць генеральної сукупності, факторних ознак на окремі одиниці генеральної сукупності, включаючи можливість впливу суттєвих сторонніх факторів на формування окремих одиниць сукупності, повноти інформації щодо генеральної сукупності.

3. Порівняння генеральної сукупності з сукупністю спостереження за попередні періоди на мікрорівні. При цьому використовувати найбільш актуальні дані спостереження.

4. Визначення статистичних одиниць, що є в генеральній сукупності, але відсутні в сукупності одиниць спостереження та навпаки. При цьому необхідно проаналізувати відповідні розбіжності. Щодо одиниць, яких немає в генеральній сукупності, слід чітко вияснити причини неузгодженості. Далі необхідно прийняти рішення, яким чином включати такі статистичні одиниці до вибіркової сукупності при проведенні державного статистичного спостереження. Зокрема, їх можна відбирати з використанням механізму простого випадкового відбору або включати всі такі одиниці без відбору при невеликій їх кількості.

5. Визначення сукупності статистичних одиниць, що є і в генеральній сукупності і в сукупності одиниць спостереження, з якою проводили співставлення, а також щодо яких є всі необхідні дані (класифікаційні ознаки, показники тощо). На основі цієї сукупності розробляється план вибірки.

6. Складання розподілу ознак за основними характеристиками сукупності та їх графічне представлення, а також кореляційний аналіз основних статистичних показників.

7. Розрахунок та оцінка статистичних характеристик сформованих розподілів: середніх значень, середніх квадратичних відхилень, коефіцієнтів варіації тощо для основних показників сукупності. Для показників, що мають ступінь варіації по одиницях сукупності вищий за встановлену норму (відповідно до теорії статистики загальноприйнята норма 30%), для відбору одиниць використовується стратифікована вибірка. Якщо варіація показників невелика, використовують простий випадковий відбір.

8. Визначення критеріїв стратифікації. Страти утворюють таким чином, щоб сукупності одиниць всередині кожної з них були якнайбільш статистично однорідними. При цьому слід враховувати, що чим ефективніша процедура стратифікації, тим точнішими будуть результати спостереження за інших рівних умов. За допомогою стратифікації отримують вищу точність результатів вибірки у порівнянні з простим випадковим відбором.

Для стратифікації, перш за все, використовують основні статистичні класифікації та ознаки (КІСЕ, КВЕД, КОАТУУ, ознака розміру одиниці тощо). Якщо ступінь варіації показника є значним, тобто сукупності одиниць у стратах залишаються неоднорідними, виникає необхідність виокремити деякі одиниці сукупності, які мали нетипові величини показника, що досліджується.

Найпоширенішим методом виявлення нетипових одиниць (аутлаєрів) є метод 3-сигма, суть якого полягає в припущенні, що значення нормально розподіленої випадкової величини показника лежать у певному інтервалі [8]. Нетипові одиниці доцільно обстежувати суцільно. Після виявлення та вилучення нетипових одиниць ступінь варіації показника у стратах повинен значно зменшитись.

9. Стратифікація сукупності.

10. Для кожної страти визначається середнє значення показника, середньоквадратичне відхилення, коефіцієнт варіації, дисперсія ознак по одиницях сукупності тощо.

11. Визначення розрізів (доменів), за якими буде поширюватися статистична інформація.

12. Визначення рівня надійності (насамперед, точності), з якою передбачається оцінювання основних показників при подальшому проведенні вибіркового обстеження.

13. Розрахунок проектного обсягу вибіркової сукупності за визначеним способом.

14. Оцінка розподілу проектного обсягу вибірки за стратами.

Розподіл проектного обсягу вибірки за стратами може бути здійснений на основі таких методів:

пропорційний розподіл – метод розподілу, за яким обсяг вибірки розподіляється пропорційно розміру страт;

рівний розподіл – розподіл, який забезпечує рівний обсяг вибірки по кожній страті;

розподіл Неймана (оптимальний розподіл) – розподіл, який забезпечує мінімальну величину дисперсії вибіркових оцінок показника по генеральній сукупності. Цей розподіл здійснюється з урахуванням розміру страт і дисперсії показника (ознаки, за якою розраховується показник) по одиницях сукупності. При цьому страти більші за розміром та/або з більшою варіацією відповідної характеристики одиниць сукупності отримують більший обсяг вибірки;

розподіл Банкайра – розподіл схожий на розподіл Неймана, однак замість дисперсії використовуються коефіцієнт варіації та додаткові ознаки.

15. З генеральної сукупності неповторним випадковим відбором вибирається задана кількість вибіркових сукупностей (реплікацій). Кількість необхідних реплікацій залежить від властивостей оціночної функції та може досягати кількох тисяч.

16. Результати вибірки розповсюджуються на генеральну сукупність.

17. Для основних показників вибіркової сукупності розраховуються похибки вибірки. Також розраховується дизайн-ефект, що визначає ефективність стратифікації.

18. Після виконання всіх розрахунків результати оформлюються у вигляді таблиць. Макети таблиць мають містити досить повну інформацію для забезпечення всебічного аналізу отриманих результатів.

19. Після аналізу результатів, отриманих за допомогою різних вибірок, сформованими різними способами та з різними критеріями стратифікації, вибирається найбільш оптимальний метод формування вибіркової сукупності з найбільш оптимальними критеріями стратифікації.

7.2. Вище описані методи охоплюють основні завдання аналізу генеральної сукупності та побудови плану вибірки. При формуванні плану вибірки методи можуть удосконалюватись, уключати в себе додатковий аналіз як характеристик самих показників, так і зв'язків між ними, а також містити більш складні методи відбору одиниць.

7.3. При впровадженні нового вибіркового спостереження доцільно провести пілотне обстеження, за результатами якого побудувати план вибірки. Також іноді використовують додаткову інформацію, наприклад дані інших спостережень, адміністративні дані тощо.

7.4. При удосконаленні вже існуючого вибіркового спостереження переглядається план вибірки у частині ефективності методу відбору, методів отримання оцінок, обсягу вибірки тощо. Отримані результати за новим планом порівнюються з результатами, отриманими за старим планом.

8. Методи редагування та зважування даних

8.1. Правила та методи редагування даних.

8.1.1. При проведенні масштабних державних статистичних спостережень, як правило, виникають помилки при внесенні респондентами даних до статистичних формулярів. Для виявлення таких помилок застосовують спеціальні процедури контролю та редагування первинних даних.

Найчастіше використовують такі види контролю первинних статистичних даних: повноти, арифметичний, логічний, перевірка правильності кодування.

Правила контролю можуть бути стохастичними (несуворі контролі) або детермінованими (суворі контролі).

Стохастичне правило – це попередження про можливу помилку.

Детерміноване правило – правило, яке вказує на наявність помилки з імовірністю одиниця.

Правила контролю також розрізняються за часом їх виконання на етапі обробки даних. Контроль даних здійснюється під час введення даних та/або після нього (первинні контролі та вторинні контролі).

8.1.2. Основними методами редагування даних є: виправлення помилок та імпутація (автоматизоване редагування).

8.2. Імпутація даних державних статистичних спостережень.

Імпутація пропущених даних або даних, що не пройшли контроль, є ефективним заходом зниження навантаження на респондентів та водночас забезпечує можливість підвищення якості статистичних даних, насамперед їх повноти [5].

8.2.1. Існує два види відсутності даних, які передбачалось отримати за результатами спостереження: часткове та повне. Імпутація застосовується у випадку часткової відсутності даних. У випадку повної відсутності даних, як правило, застосовують метод коригування статистичних ваг.

8.2.2. Доцільно виділити два основні підходи до обробки частково відсутніх даних.

Перший підхід (ігнорування) – часткова відсутність даних ігнорується. Такий підхід називається методом аналізу повного змісту.

Другий підхід – застосування методів статистичної оцінки відсутніх даних та їх умовне нарахування, тобто імпутація.

8.2.3. Основні методи імпутації:

метод середньої або моди. Він полягає в присвоєнні відсутньому значенню певної характеристики одиниці спостереження середнього або модального значення, визначеного за всіма одиницями сукупності або за одиницями підсукупності, до якої належить одиниця з відсутніми даними;

метод «хот-дек». Метод імпутації, який передбачає вибір даних, що імпутуються для кожного пропущеного значення показника з урахуванням розподілу ознаки і ступеня схожості одиниць вибірки за класифікаційними ознаками.

8.3. Зважування даних державних статистичних спостережень.

8.3.1. У ході вибіркового статистичного спостереження статистики підприємств отримують дані тільки для частини генеральної сукупності – для вибіркової сукупності одиниць спостереження.

Метод, що дає можливість розповсюдити інформацію по вибірковій сукупності на генеральну, називається статистичним зважуванням. Статистичне зважування є складовою процесу оцінювання показників за результатами вибіркового спостереження. Головну роль в процесі розповсюдження характеристик вибірки на генеральну сукупність відіграють імовірності включення одиниць до вибірки, оскільки у відповідності з процедурою формування вибірки кожна одиниця вибіркової сукупності представляє певну кількість одиниць генеральної сукупності (ця кількість визначає величину початкової (базової) статистичної ваги одиниці спостереження, що є оберненою до ймовірності включення одиниці до вибірки). При статистичному зважуванні важливе значення має також наявність даних щодо кожної одиниці з вибірки або у разі їх повної відсутності – інформація щодо причини невідповіді. Така інформація дає можливість скоригувати статистичні ваги для врахування при розповсюдженні даних на генеральну сукупність поряд з імовірністю включення одиниць до вибірки також імовірності отримання від них інформації.

8.3.2. Для здійснення оцінок статистичних показників за результатами державних статистичних спостережень отримані дані зважуються. Це забезпечує:

- врахування фактичних імовірностей включення одиниць генеральної сукупності до вибірки;

- врахування фактичних рівнів участі одиниць спостереження в вибіркового статистичному спостереженні;

- зменшення впливу недосконалостей основи вибірки та інших окремих невібіркових похибок на результати обстеження;

- зіставність статистичних даних з даними інших статистичних спостережень.

8.3.3. Зважування здійснюється на основі розрахунку системи статистичних ваг. Розрахунок системи ваг включає такі основні етапи:

- розрахунок базової ваги одиниць вибірки;

- коригування базової ваги для врахування ступеня участі відібраних одиниць у державному статистичному спостереженні (коригування на невідповіді);

- коригування ваг для узгодження результатів вибіркового статистичного спостереження з наявною зовнішньою інформацією, при застосуванні оцінок за моделями тощо – калібрація системи статистичних ваг;

- контроль якості системи ваг.

9. Рекомендовані стандартні програмні продукти для проведення державних статистичних спостережень

Відповідно до Критеріїв, за якими державні статистичні спостереження відносяться до статистики підприємств, для побудови вибірки та оцінки отриманих результатів вибірових спостережень статистики підприємств мають застосовувати єдині програмні платформи для обробки даних.

На цей час для цілей державних статистичних спостережень використовують PASW Statistics (SPSS). У цьому програмному забезпеченні передбачена можливість використання практично всіх необхідних процедур і методів відбору одиниць, обробки даних та зважування вибірового спостереження підприємств. Додатково можуть бути використані допоміжні програми, адаптовані до PASW Statistics (SPSS), g-Calib для проведення калібрування ваг та WesVAR для отримання більш точних оцінок надійності результатів вибірового статистичного спостереження при його складному дизайні [7].

Список використаних джерел

1. Закон України "Про офіційну статистику" від 16 серпня 2022 року № 2524-IX.

URL: <https://zakon.rada.gov.ua/laws/show/2524-20#Text>

2. Вибіркове спостереження. Термінологічний словник / О. О. Васечко, О. І. Черняк, Є. М. Жуйкова та інші.- К.: Державний комітет статистики України, 2004. – 140 с.

3. Проблеми статистичного зважування вибірових даних [Саріогло В. Г.]: Монографія. – К. ІВЦ Держкомстату України, 2005. – 264 с.

4. Survey sampling reference guidelines, 2008 edition. Introduction to sample design and estimation techniques.

URL: <https://ec.europa.eu/eurostat/documents/3859598/5901961/KS-RA-08-003-EN.PDF/833f7740-0589-47e1-99a5-c14878a2c1a8?t=1414781399000>".

Порядок здійснення робіт із застосування вибіркового методу обстеження в державних статистичних спостереженнях статистики підприємств

1. За результатами проведення інвентаризації державних статистичних спостережень самостійний структурний підрозділ Держстату, відповідальний за координацію робіт з питань статистичної методології, визначає перелік державних статистичних спостережень, для яких надані рекомендації застосування вибіркового методу обстеження (переведення суцільного спостереження або спостереження основного масиву на вибірковий метод обстеження, впровадження нового вибіркового статистичного спостереження, удосконалення вже існуючого вибіркового статистичного спостереження).

2. Керівник державного статистичного спостереження (далі – Керівник ДСС) за участю фахівців самостійного структурного підрозділу Держстату, відповідальних за координацію робіт з питань статистичної методології, відповідно до компетенції за визначеним переліком здійснює аналіз державного статистичного спостереження та викладає його результати з висновками та пропозиціями щодо доцільності/недоцільності застосування вибіркового методу обстеження за структурою, наведеною у додатку 2 до цих Методологічних положень.

3. Керівник ДСС подає заповнену форму з висновками та пропозиціями щодо доцільності/недоцільності застосування вибіркового методу обстеження та можливих виконавців побудови плану вибірки для державного статистичного спостереження на розгляд Комісії з питань методології та звітної-статистичної документації (далі – Комісія).

4. За результатами розгляду Комісія виносить рішення щодо доцільності/недоцільності застосування вибіркового методу обстеження в державному статистичному спостереженні та виконавців побудови плану вибірки за визначеними у звіті пропозиціями.

5. За результатами рішення Комісії у разі схвалення доцільності застосування вибіркового методу обстеження керівник ДСС складає План-графік робіт щодо застосування вибіркового методу обстеження в державному статистичному спостереженні (далі – План-графік), у якому зазначаються основні заходи, їх виконавці та строки виконання. План-графік погоджується з виконавцями побудови плану вибірки і затверджується першим заступником Голови Держстату України. Самостійний структурний підрозділ Держстату, відповідальний за координацію робіт з питань статистичної методології, здійснює координацію та моніторинг виконання заходів Плану-графіка.

6. У разі, якщо виконавцем розробки плану вибірки визначено самостійний структурний підрозділ Держстату, відповідальний за координацію робіт з питань статистичної методології, керівник ДСС за його участю складає замовлення на розробку плану вибірки (форму замовлення наведено у додатку 3).

7. Самостійний структурний підрозділ Держстату, відповідальний за координацію робіт з питань статистичної методології, здійснює розробку плану вибірки, керуючись цими Методологічними положеннями, у строки, визначені Планом-графіком.

8. Після розробки плану вибірки самостійний структурний підрозділ Держстату, відповідальний за координацію робіт з питань статистичної методології, складає звіт, у якому описує методи, що були застосовані при розробці плану вибірки, а також отримані результати симуляційного дослідження і надає звіт керівнику ДСС.

9. Керівник ДСС вивчає звіт на відповідність вимогам, вказаним у замовленні на розробку плану вибірки та/або розробку програмного забезпечення.

10. У разі, якщо отримані результати після розробки плану вибірки відповідають вимогам, вказаним у замовленні, керівник ДСС проводить апробацію проведення державного статистичного спостереження із застосуванням вибіркового методу за участю самостійного структурного підрозділу Держстату, відповідального за координацію робіт з питань статистичної методології у частині формування вибіркової сукупності, зважування та розрахунку похибок вибірки.

11. За результатами проведення апробації державного статистичного спостереження із використанням вибіркового методу обстеження самостійний структурний підрозділ Держстату, відповідальний за координацію робіт з питань статистичної методології, у разі потреби здійснює доопрацювання плану вибірки.

12. Після доопрацювання плану вибірки керівник ДСС за участю фахівців самостійного структурного підрозділу Держстату, відповідального за координацію робіт з питань статистичної методології, складає та вносить на розгляд Комісії проект методологічних положень щодо проведення державного статистичного спостереження із застосуванням вибіркового методу обстеження.

13. Після затвердження методологічних положень керівник ДСС забезпечує проведення державного статистичного спостереження із використанням вибіркового методу обстеження відповідно до цих положень.

14. Самостійний структурний підрозділ Держстату, відповідальний за координацію робіт з питань статистичної методології, здійснює заходи із передачі методологічно-практичних навичок проведення державного статистичного спостереження із застосування вибіркового методу (формування вибіркової сукупності, зважування, розрахунку похибок вибірки тощо), а також програмного забезпечення із формування вибіркової сукупності, зважування та розрахунку похибок вибірки керівнику ДСС.

15. У разі, якщо отримані результати після розробки плану вибірки не відповідають вимогам, вказаним у замовленні, керівник ДСС змінює вимоги у замовленні та надає його самостійному структурному підрозділу Держстату, відповідальному за координацію робіт з питань статистичної методології, або здійснює заходи із залучення науково-дослідних установ, інших організацій

та\або міжнародних експертів в установленому порядку для доопрацювання плану вибірки.

Результати аналізу державного статистичного спостереження з висновками та пропозиціями щодо доцільності / недоцільності використання в ньому вибіркового методу обстеження

(назва та код ДСС за планом державних статистичних спостережень)

(назва та код самостійного структурного підрозділу Держстату, уповноваженого на проведення ДСС)

- 1. Відповідність:**
 - 1.1. Мета обстеження _____
 - 1.2. Основні показники _____
 - 1.3. Статистична одиниця _____
 - 1.4. Період статистичного спостереження _____
 - 1.5. Вид статистичного спостереження _____
 - 1.6. Кількість одиниць в генеральній сукупності _____
 - 1.7. Кількість одиниць в сукупності, що обстежується _____
 - 1.8. Рівень деталізації показника за основними класифікаційними ознаками _____

- 2. Своєчасність та пунктуальність:**
 - 2.1. Тривалість збору даних _____
 - 2.2. Тривалість обробки даних _____
 - 2.3. Термін оприлюднення статистичної інформації _____

- 3. Оцінка потреб та очікувань користувача:**
 - 3.1. Перелік основних користувачів (у тому числі інші державні статистичні спостереження та/або національні рахунки) _____
 - 3.2. Вимоги користувачів щодо статистичної інформації, що надається _____

- 4. Вимоги щодо точності оцінювання показників у разі застосування вибіркового методу:**

- 5. Вимоги щодо програмного забезпечення** _____

- 6. Інші особливі вимоги щодо застосування вибіркового методу обстеження:** _____

- 7. Виконавець розробки плану вибірки** _____

- 8. Висновки та пропозиції щодо доцільності / недоцільності застосування вибіркового методу обстеження у державному статистичному спостереженні.**

Керівник ДСС

(дата)

_____ (підпис)

_____ (ініціали (ініціал імені), прізвище)

ЗАМОВЛЕННЯ № _____
на розробку плану вибірки та/або розробку програмного забезпечення

(назва та код ДСС за планом державних статистичних спостережень)

(назва та код самостійного структурного підрозділу Держстату, уповноваженого на проведення ДСС)

1. Основні показники, за якими розробляється план вибірки:

2. Критерії для формування генеральної сукупності одиниць:

інституційний сектор економіки, який обстежується _____
тип статистичної одиниці _____
організаційно-правові форми господарювання (за КОПФГ) _____
вид економічної діяльності (за КВЕД) _____
розмір підприємства _____
додаткові критерії¹: _____

3. Вимоги щодо точності оцінювання показників

4. Вимоги щодо програмного забезпечення в частині формування вибіркової сукупності, зважування, розрахунку похибок вибірки:

5. Інші особливі вимоги щодо застосування вибіркового методу обстеження:

Дата подання замовлення _____

Дата отримання замовлення _____

Керівник ДСС _____

(дата)

(підпис)

(ініціали (ініціал імені), прізвище)

¹ У межах інформаційного фонду РСО