

А. М. ЄРІНА, З. О. ПАЛ'ЯН

СТАТИСТИЧНІ СПОСТЕРЕЖЕННЯ:

переписи,
моніторинги,
вибіркові
обстеження

Навчальний
посібник



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

А. М. Єріна, З. О. Пальян

**СТАТИСТИЧНІ СПОСТЕРЕЖЕННЯ:
ПЕРЕПИСИ, МОНІТОРИНГИ,
ВИБІРКОВІ ОБСТЕЖЕННЯ**

Навчальний посібник

КИЇВ 2019

УДК 331.21(075.8)

Е 78

Рецензенти:

О. М. Гладун, д.е.н., чл.-кор. НАН України
Інститут демографії та соціальних досліджень ім. М.В. Птухи НАН України»
В. Г. Швець, д.е.н., проф.
Київський національний університет імені Тараса Шевченка

*Рекомендовано до друку
вченою радою економічного факультету
Київського національного університету імені Тараса Шевченка
Протокол № 11 від 14 травня 2019 року*

С 78 **Статистичні спостереження: переписи, моніторинги, вибіркові обстеження : навчальний посібник / [А. М. Єріна, З. О. Пальян]. — К. :, 2019 — 308, [4] с.**
ISBN

У навчальному посібнику розглядаються основні види, способи, програмно-методологічні й організаційні питання статистичного спостереження в контексті інформаційного забезпечення системи управління і наукових досліджень. Висвітлюються питання сучасної організації переписів і моніторингів, а також формування інтегрованих статистичних банків даних та ведення статистичних реєстрів у комбінації з іншими формами спостереження. Значна увага приділена вибірковому методу, зокрема: схемам формування вибіркових сукупностей, обґрунтуванню обсягів і дизайну вибірки, принципам забезпечення репрезентативності й надійності вибіркових оцінок. Аналітичні можливості та межі застосування конкретних методів статистичного спостереження ілюструються на прикладах, різних за соціально-економічним змістом та інформаційною базою.

Для студентів, аспірантів, молодих науковців і практиків — для усіх, хто прагне опанувати сучасні методи статистичного спостереження, зокрема, методи вибіркових обстежень.

УДК 331.21(075.8)

ISBN

© А. М. Єріна, З. О. Пальян, 2019

© КНУ імені Тараса Шевченка,
ВПЦ «Київський університет», 2019

ПЕРЕДМОВА

На етапі переходу до інформаційного суспільства зростає попит на статистичну інформацію, змінюється характер вимог до неї — від фіксації певних фактів до

всестороннього аналізу процесів, що відбуваються в суспільстві, та прийняття обґрунтованих управлінських рішень. Для забезпечення потреб управління в аналітичній і прогностичній інформації щодо важливих сфер діяльності суспільства у практику статистичних досліджень запроваджуються нові, альтернативні традиційній звітності, більш ефективні статистичні й інформаційні технології, форми і способи статистичного спостереження.

Нинішня система державних статистичних спостережень здебільшого організується у формі спеціальних, у тому числі несудових обстежень, або як комбінація кількох видів, форм і способів спостереження. Тому вкрай важливо не лише знати сфери застосування різних статистичних спостережень, а й уміти розробляти програмно-методологічні й організаційні питання плану конкретних обстежень з огляду на природу соціально-економічних явищ та можливості інформаційного забезпечення. Саме такі знання і навички допомагає сформувати навчальний посібник «Статистичні спостереження: переписи, моніторинги, вибіркові обстеження».

У посібнику чільне місце відведено питанням сучасної організації переписів, що особливо актуально у період підготовки країн світу та України до проведення чергового раунду переписів населення та житлового фонду. Неабияку увагу приділено проблемі формування інтегрованих статистичних банків даних та ведення статистичних реєстрів у комбінації з іншими формами спостереження. Окремий розділ присвячено особливостям організації в різних сферах життя відносно нового виду спостереження — статистичного моніторингу (екологічного, соціального, ділової активності підприємств тощо).

У світі невпинно розширюється практика вибіркового обстеження з надзвичайно широкої проблематики. У державній статистиці традиційно найбільш масштабним є обстеження домогосподарств, які дають змогу отримати інформацію щодо зайнятості населення, доходів і витрат, споживання

матеріальних благ і пос-

луг, житлових умов та інших характеристик життєвого рівня населення. Вельми швидких обертів набирають вибіркові обстеження у сфері комерції та бізнесу, при вивченні сегментації ринку, ефективності реклами і споживчих реакцій на окремі товари, в інших аспектах підприємницької діяльності. Розширюється сфера використання вибіркових опитувань з метою аналізу політичної ситуації, соціального напруження тощо.

Зібрати дані не про всю сукупність, а лише про її частини, простіше, швидше, дешевше, але дібрана частина репрезентуватиме ціле з прийнятним ступенем точності лише за певних умов. У посібнику розглянуто накопичений міжнародний і вітчизняний досвід організації вибіркових обстежень, методологічні принципи і способи формування вибіркових сукупностей, правила і методи оцінювання надійності результатів вибіркового обстеження, способи поширення вибіркових даних на генеральну сукупність. Усі вони спрямовані на забезпечення репрезентативності результатів обстежень, їхню ефективність проілюстровано на конкретних прикладах соціально-економічного змісту.

Структура посібника підпорядкована навчальним програмам магістерських курсів «Статистичне спостереження» та «Вибіркові спостереження», мета яких — набуття теоретичних знань і практичних навичок у підготовці й організації сучасних форм і видів статистичного спостереження, веденні статистичних реєстрів, формуванні вибіркових сукупностей, результати обстеження яких можуть забезпечити необхідну точність даних із мінімальними витратами.

Для систематизації та кращого засвоєння навчального матеріалу, а також набуття практичних навичок статистичного спостереження до кожного розділу додаються вправи і запитання для самоконтролю

Розділи 1–4 написані З.О. Пальян, розділи 5–10 — А.М. Єріною.

Розділ 1

СТАТИСТИЧНІ СПОСТЕРЕЖЕННЯ ЯК МЕТОД ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ

Статистичне спостереження — перший етап статистичного дослідження

Джерела статистичних даних, вимоги, яким вони мають відповідати

Програмно-методологічні питання статистичного спостереження. Помилки спостереження

Організаційні питання статистичного спостереження

Після вивчення цього розділу студент має

- *знати*: сутність статистичного спостереження як методу збирання масових даних; особливості статистичної інформації; вимоги до статистичних даних; риси сучасного характеру та способи використання статистичних даних;
- *уміти*: користуватися різними джерелами статистичних даних; визначати мету, об'єкт, одиницю спостереження та одиницю сукупності; розробляти план і програму спостереження; вирішувати організаційні питання спостереження; виконувати логічний та арифметичний контроль даних.

Статистичне спостереження — перший етап статистичного дослідження

Якщо запитати пересічну особу: «Чим займається статистика?», то найімовірніше отримаємо відповідь: «Статистика збирає, обробляє та аналізує інформацію». Отже, кожне статистичне дослідження розпочинається із пошуку, збирання, реєстрації та нагромадження даних про різноманітні сфери суспільного життя, про явища і процеси живої природи.

Утім, завдання інформаційного забезпечення є першочерговим не лише для статистичного дослідження, а й для будь-якої сфери діяльності. У широкому розумінні *інформаційне забезпечення* — це сукупність відомостей про явища та процеси життя суспільства, які зорієнтовані на певних користувачів. Тож інформація — це продукт збирання та оброблення даних, який має від-

повідне аналітичне призначення. Зрозуміло, що інформацію отримують не заради «інформації», тобто лише самого процесу її пошуку. Кінцева мета — це підготовка висновків та прийняття управлінських рішень, спираючись на якісну інформацію. Так, відомості про чисельність і склад населення за віком працездатності є підґрунтям для формування державної стратегії зайнятості і працевлаштування, розвитку соціальної інфраструктури та соціальних послуг тощо. Приміром, бюджети домогосподарств слугують підставою для визначення реального споживчого кошика в країні, розроблення програми соціального захисту малозабезпечених верств населення, визначення механізму індексації доходів населення. Інформація про ринок житлової нерухомості використовується не лише для визначення масштабів забудови територій, формування маркетингової і цінової політики. Визначення кола партнерів та конкурентів на ринку допомагає запобігти банкрутствам, скоординувати роботу окремих комерційних і державних структур.

З точки зору технології збирання й опрацювання статистичних даних *інформаційне забезпечення* складається з чотирьох взаємопов'язаних елементів: статистичних формулярів; нормативних документів; методичних матеріалів, що стосуються: процедури збирання даних, порядку класифікації та кодування, обсягу охоплення сукупності, алгоритму розрахунку показників, форми організації інформації; статистичних даних.

За сучасних умов формування ринкових відносин дещо змінюються пріоритети в характері інформаційного забезпечення статистичними даними. Першочергове завдання системи державної статистики полягає не лише у задоволенні потреб користувачів інформаційного продукту. Акцент робиться на інформаційну підтримку прийняття управлінських рішень в усіх ланках. Зокрема, поглиблюються аналітичні можливості статистичної інформації, створюються моделі для короткострокових прогнозів соціально-економічних явищ. Розширюється коло показників, за якими здійснюються міжнародні порівняння. Відбувається гармонізація національної системи статистичних показників соціально-економічного розвитку країни із міжнародними стандартами. Зокрема, чинна в Україні Система національних рахунків (СНР 2008) побудована за методологічними принципами міжнародних стандартів Європейської системи національних та регіональних рахунків 2010 року (ESA 2010) з урахуванням особливостей української економіки й законодавства.

Паралельно із традиційною національною методикою обчислення показників використовується міжнародна, а разом і критерії їхнього оцінювання. Прикладом може бути визначення рівня зайнятості та безробіття в країні: за методологією Міжнародної організації праці (на підставі вибіркового обстеження домогосподарств) та рівня зареєстрованого безробіття за даними Державної служби зайнятості України.

Проте чинні умови господарювання, зокрема наявність недержавних структур, висувають особливі вимоги до інформаційного забезпечення. Зрозуміло, що отримання даних про діяльність таких форм підприємництва, як корпорації, холдингові компанії чи фізичних осіб-підприємців не може відбуватися в однаковий спосіб. Це стосується і реєстрації результатів діяльності різноманітних некомерційних організацій: виробничих і торговельних асоціацій, консалтингових фірм, громадських організацій та фондів тощо. Доступ до такої інформації нерідко ускладнюється через наявність перешкод на кшталт: комерційна таємниця, невідповідність юридичної та фактичної адрес, а також зареєстрованого і фактичного виду економічної діяльності. Нерідко одиниці господарювання, особливо малі підприємства, відмовляються від надання інформації, посилаючись на надмірне звітне навантаження.

З огляду на це органами державної статистики розробляються й удосконалюються методи і прийоми збирання різнобічних статистичних даних. Так, згідно із європейськими нормами і правилами, державні органи статистики мають забезпечити конфіденційність статистичних даних респондентів. З цією метою здійснюється деперсоніфікація первинних даних спостереження. *Деперсоніфікація* — знеособлення зареєстрованої і обробленої статистичної інформації, наданої одиницями спостереження. В Україні цю сферу контролюють два Закони: «Про доступ до публічної інформації» та «Про захист персональних даних». Законодавчо визначено, яка інформація належить до так званих відкритих даних (“open data”), щодо якої розпорядники інформації зобов’язані надавати дані на вимогу, оприлюднювати та регулярно оновлювати її на Єдиному державному веб-порталі відкритих даних. Прикладом деперсоніфікованих масивів даних може бути Банк персональних даних учасників зовнішнього незалежного оцінювання знань з предметів (ЗНО) Міністерства освіти України.

Деперсоніфікація здійснюється кількома способами. Один з найпоширеніших — вилучення прямих персональних реквізитів та

їхнє перекодування або зміна унормованого порядку розташування одиниць у переліку чи списку. Іноді застосовують комбінацію ознак, яка внеможливилоє ідентифікацію одиниць. Анонімність первинних даних забезпечується також заміною індивідуальних значень показників, що можуть використовуватися в якості ключа-ідентифікатора, на сумарне, середнє за сукупністю значення.

Задля зменшення звітнього навантаження суб'єктів господарювання оптимізується періодичність надання звітної інформації, частіше використовується адміністративна інформація, що міститься у Державних галузевих чи цільових реєстрах (Єдиний державний реєстр платників податків, Єдиний державний демографічний реєстр тощо). Ширшого впровадження набувають несутільні вибіркові обстеження. В Україні запроваджено щорічні (внутрішньорічні) інтегровані вибіркові обстеження діяльності підприємств, їхньої ділової активності тощо.

Збирання надійних даних пов'язане з ретельним науковим обґрунтуванням мети та способу отримання інформації, її подальшого використання. Тому процес формування якісної інформаційної бази потребує чіткої спланованості на першому етапі статистичного дослідження, яким є статистичне спостереження.

Статистичне спостереження — це спланована, науково організована реєстрація масових даних про будь-які соціально-економічні явища та процеси. Від інших методів збирання даних статистичне спостереження відрізняється характером і масовістю даних та способами їх отримання. Крім безпосередньої реєстрації (вимірювання, підрахунок, оцінювання), широко застосовується вивчення громадської думки на підставі опитувань, референдумів.

Зауважимо, що статистичні спостереження охоплюють дані різного ступеня реєстрації, тому вони можуть бути первинними або вторинними. *Первинне спостереження* — це реєстрація вихідних даних, що надходять від об'єкта, який їх продукує. Прикладами можуть бути поточний облік кількості зареєстрованих шлюбних пар і розлучень або звіти суб'єктів господарювання про результати виробничої та фінансової діяльності, чи опитування населення стосовно довіри до органів державного управління, силових структур.

Вторинне спостереження — це збирання раніше зареєстрованих та оброблених даних, наприклад матеріалів фінансових звітів підприємств, обліку запасів палива в регіоні, результатів зовнішньоекономічної діяльності країни.

Статистичне спостереження здійснюється в три етапи: підготовка спостереження; реєстрація статистичних даних; формування бази даних.

На підготовчому етапі вирішують, які дані збиратимуться, в який спосіб реєструватимуться і подаватимуться, які джерела інформації використовуватимуться.

Джерела статистичних даних, вимоги, яким вони мають відповідати

У глобальному інформаційному просторі статистичні дані посідають окреме місце за своїм характером, способом отримання та метою використання. Побуває думка, що статистичні дані неточні, приблизні, оманливі або відверто брехливі. Хто тільки не оперує статистичною інформацією, аби довести чи заперечити певні судження, викрити або виправдати певні дії. Статистичними даними користуються пересічні особи та фахівці різного спрямування: практики, науковці, політики, управлінці. При цьому керуються різною метою. Ці дані застосовують для формалізованого оцінювання стану явищ і перебігу процесів, для порівняльного аналізу й узагальненої характеристики процесів чи об'єктів. Статистичні дані слугують також як прогнозні, очікувані оцінки задля передбачення чи запобігання бажаним або небажаним змінам. Ними користуються для більшої переконливості тверджень, думок, висновків та пропозицій.

Особливістю статистичної інформації є її кількісна визначеність. Парадоксальність полягає в тому, що статистичні дані дають кількісну, цифрову оцінку, в тому числі й тим явищам, станам і процесам, які не підлягають безпосередньому вимірюванню. Пригадаймо принаймні, яким чином оцінюється рівень набутих знань і навичок при навчанні, професійної та виконавчої майстерності, якості надання клієнтам сервісних послуг. У цьому разі вимірювання зводиться до оцифрування окремих градацій якості — «відмінно», «добре» тощо. За таким принципом здійснюється статистичне оцінювання значно складніших суспільних явищ і процесів. Скажімо, кількісно описують фінансову стабільність клієнтів банків або надійність компаній партнерів з бізнесу. За допомогою системи статистичних показників та індикаторів «вимірюють» різноманітні аспекти розвитку країн світу та будують відповідні рейтинги. До прикладу: рейтинги економічного розвитку, соціальної,

екологічної й інформаційної безпеки, за глобальним індексом миролюбності, людського розвитку, за індексом щастя.

Статистичні дані описують явища та процеси, які мають стохастичну природу. Тобто їм властиве формуватися та проявлятися лише в масі випадків. Скажімо, на запитання, кого народжується більше — хлопчиків чи дівчаток — кожен може відповісти по-різному, з огляду на власні спостереження або судячи з епізодичних повідомлень мас-медіа. Але за достатньо великої кількості народжень у регіоні чи країні завжди діє закономірність переважання чоловічої статі серед новонароджених.

Різноманітність статистичних даних та сфер їх використання зумовлює різні джерела цього інформаційного продукту. Залежно від того, в якій формі подаються та яким чином збираються статистичні дані, можна класифікувати їхні джерела (рис. 1.1).



Рис. 1.1. Класифікація джерел статистичних даних

Очевидно, що основним джерелом статистичних даних є статистична звітність та матеріали первинного, в тому числі бухгалтерського обліку, якщо йдеться про результати економічної діяльності суб'єктів господарювання. Інформацію щодо соціально-політичних, демографічних, екологічних та інших процесів збирають шляхом первинного обліку, реєстрації відповідних подій, моніторингу їхнього перебігу. Широко використовуються дані опитувань, вибіркові обстеження, а в разі суцільних спостережень — переписи.

Доступ до вже зібраної, систематизованої та опрацьованої статистичної інформації здійснюється через офіційні документи, зокрема зведені статистичні таблиці, що готуються органами Державної статистики. Широкому загалу доступні загальні та спеціалізовані публікації, а саме: статистичні збірники, бюлетені, експрес-випуски, аналітичні звіти та огляди, які зазвичай містяться в окремих брошурах, фахових журналах тощо.

Похідну (вторинну) статистичну інформацію користувач може знайти також у профільних наукових публікаціях або офіційних виданнях державних структур, але з обов'язковим посиленням на першоджерело даних, а саме на Державну службу статистики чи спеціально організоване обстеження, що проведене міністерством, відомством, установою, фірмою, агенцією або фізичною особою, скажімо, науковцем чи дослідником.

Кардинальні зміни в систему формування та доступу до статистичних даних внесли електронні засоби збирання, опрацювання та подання статистичної інформації. Загальнодоступним джерелом статистичних даних є офіційні Web-сайти, мережі internet та різні соціальні мережі. У теперішній час це найоперативніший спосіб як отримання даних, так і їх оприлюднення. Але відкритість сайтів та соцмереж у плані подання інформації не завжди гарантує якість розміщених даних, зокрема їхню коректність та правдивість.

Важливим напрямом роботи Державної служби статистики України є широке запровадження електронної звітності для всіх суб'єктів господарювання, що спрощуватиме процедуру збирання звітних даних і сприятиме гнучкому доступу до зведеної та систематизованої інформації.

Швидкому реагуванню на потреби та дії суб'єктів економічної діяльності або населення з питань соціально-економічного, політичного чи мотиваційного характеру сприяє запровадження інте-

рактивної мережі запиту та накопичення статистичних даних. Уже зараз широко застосовується формат і спосіб поширення інформації, що функціонує в технологічному режимі «клієнт — сервер». Інтеграція регіональних мереж документообігу в єдину систему загальнодержавних банків статистичних даних забезпечує не лише доступність даних, а й гарантує їхню відповідну якість — зіставність, тобто порівнянність за методикою їх збирання, опрацювання та аналізу.

Узагальнюючи, можна окреслити особливі риси сучасного характеру та способу використання статистичних даних:

- уніфікованість за формою подання, збереження та пошуку;
- глобалізація та інтеграція інформації в архівних банках даних;
- простота перенесення даних з одних носіїв на інші;
- діалоговий режим запиту, отримання даних та прийняття оперативних управлінських рішень;
- складність захисту даних від зовнішнього втручання.

Користуючись статистичними даними, ми лише зрідка замислюємося над тим, а якими вони мають бути, яким параметрам відповідати. З-поміж різноманітних інформаційних носіїв статистичні дані вирізняються *трьома основними властивостями*: це масовість; системність; кількісна визначеність.

Статистичні дані — це масові системні кількісні характеристики соціально-економічних явищ і процесів. Саме масовість відрізняє статистичні дані від інших, оскільки завдяки переходу від окремих фактів до масових можна визначити загальну закономірність, позбавлену впливу випадкових причин. Ступінь масовості залежить від рівня узагальнення досліджуваних явищ. На макрорівні збирають дані про явища та процеси загальнодержавного характеру. Це відомості про економічний потенціал держави, її платоспроможність та міжнародну інвестиційну позицію у світовому середовищі, розмір її фінансових, природних та людських ресурсів, масштаби експортно-імпорتنих операцій, інтенсивність інноваційних процесів, рівень соціальної напруги та безпеки. На мікрорівні дані також масові, але іншого гатунку. Матеріали, що їх збирають окремі міністерства, відомства, підприємства, установи, товариства, фірми та агентства, характеризуються певною фрагментарністю. Прикладом є дані про кількість зайнятих працівників на певному підприємстві, результати його виробничої діяльності та фінансові результати до оподатку-

вання, або розмір посівних площ, валовий збір та урожайність сільськогосподарських культур у домогосподарствах сільської місцевості тощо.

Статистичні дані обов'язково мають кількісну визначеність, завдяки чому підлягають накопиченню, зведенню та узагальненню. Останнє можливе лише за умов їхньої системності. Тож статистичні дані — це не «будь-які цифри під рукою», а лише факти, що узгоджуються, порівнюються і в сукупності відбивають певний бік явища чи процесу. Так, аналізуючи інвестиційні процеси в країні, слід зважати не лише на розмір інвестицій, а й на їхню структуру: внутрішні та іноземні, а також на джерела фінансування та напрями інвестування за видами економічної діяльності.

Зрозуміло, що від якості даних статистичного спостереження залежать результати подальшого дослідження і, найголовніше, прийняття і впровадження управлінських рішень. Тому вони мають відповідати певним вимогам.

Перша — це *достовірність даних*, тобто їхня відповідність реальному стану. На жаль, ця вимога іноді не додержується як на мікро-, так і на макрорівні. Прикладом можуть слугувати матеріали переписів населення на території України в період громадянської та Другої світової воєн. Вони не врахували значних втрат населення внаслідок репресій, розкуркулювання та голодомору. На мікрорівні типовими є приховування частини доходів від оподаткування як юридичними, так і фізичними особами. Або інший приклад: перебільшення розміру кредитних ресурсів комерційних банків для забезпечення нормативного мінімуму, встановленого Нацбанком України. Досягти достатньої достовірності можна, з одного боку, усуненням умов для таких викривлень (приміром, удосконалення податкової системи або зміни кредитної політики Нацбанку стосовно комерційних банків), а з іншого — застосуванням більш чутливої системи показників, які мають стати справжніми індикаторами.

Друга вимога — це *повнота даних* як за їхнім обсягом, так і за суттю. Наприклад, досліджувати ринок покупців певного товару слід з огляду не лише на ті верстви населення, на які він розрахований, а й на категорії споживачів, які можуть придбати товар для своїх дітей, родичів, друзів. Щодо повноти за суттю, то вона забезпечується системним добором кількох взаємозалежних ознак явища. А саме, виконуючи сегментацію ринку товарів чи сервісних послуг, слід проаналізувати статеву-вікову структуру

потенційних споживачів, поцікавитися їхніми потребами, вимогами, врахувати платоспроможність. Водночас необхідно вивчати і враховувати стратегію конкурентів на ринку.

Третя вимога — це *своєчасність даних*. Інформація має дійти до користувача, перш ніж застаріє, інакше вона втрачає корисність. Приміром, дані котирування цін на біржах або валютних курсів мають надходити до користувача мірою їх виникнення та реєстрації; цінність даних вибіркових обстежень домогосподарств щодо самооцінки стану здоров'я та причин незадоволеності наданням медичних послуг втрачається з часом.

Четверта вимога — це *порівнянність даних у часі або в просторі*. Дані можуть бути не порівнянними за складом сукупності. Наприклад, нестійкість притаманна структурі інвесторів за напрямами вкладання коштів, або покупців і продавців на будь-якому ринку, в тому числі житлової й нежитлової нерухомості. Гострою є проблема *порівнянності за одиницями вимірювання*, особливо це стосується вартісних показників за умов нестабільності грошової одиниці, зростання інфляції або внаслідок грошової реформи.

Із переходом національної статистики на міжнародні стандарти слід дотримуватися *порівнянної методики збирання даних та обчислення статистичних показників*. Так, програмою першого Всеукраїнського перепису населення затверджено одиницю спостереження, що використовується у світовій практиці, а саме: «домогосподарство» замість «родина», як це було в попередніх переписах. «Домогосподарство» — це сукупність осіб, що мешкають разом і ведуть спільне господарство, повністю або частково об'єднують та витрачають кошти. Очевидно, ця категорія ширша від попередньої, оскільки не обмежується лише родинними зв'язками та кровною спорідненістю. Проблема порівнянності методик на макроекономічному рівні частково розв'язується із переходом на систему національних рахунків.

Важливою є порівнянність даних за територіальною належністю одиниць унаслідок відповідних змін, наприклад, адміністративно-територіальних змін внаслідок анексії та військової окупації частини території або розширення меж обласних центрів, зокрема столичного регіону за рахунок прилеглих територій.

П'ята вимога — *доступність даних*. Із утворенням нових форм господарювання питання доступності даних особливо загострюється. Ці дані іноді важко зібрати, особливо отримати від

недержавних структур. Низький рівень відповідальності їхніх керівників при поданні звітів або непідзвітність деякої їхньої частини призводить до порушення вимог повноти та своєчасності реєстрації даних. Останнім часом спостерігається деяке поліпшення звітної дисципліни суб'єктів господарювання різних форм власності завдяки зменшенню звітного навантаження на облікові одиниці, використанню адміністративних даних, запровадженню електронної форми звітності в системі органів державної статистики. Як уже зазначалося, за сучасних умов інформаційного забезпечення, зокрема мережі internet, поступово зникає проблема доступу до готової зведеної статистичної інформації. Цьому сприяють створені Держстатом України офіційні електронні банки даних, що містять архівну інформацію і дозволяють здійснювати моніторинг соціально-економічних процесів.

Одержання якісних статистичних даних значною мірою залежить від того, на якому рівні збирається відповідна інформація. В Україні діють дві системи: централізована (загальнодержавна) та децентралізована (відомча, окремих економічних структур). Звісно, централізована система збирання має ширші можливості для якісного спостереження: наукову методіку, кваліфіковані кадри, технічне забезпечення тощо. Проте децентралізована система є оперативнішою завдяки меншій тривалості часу між збиранням даних і використанням готової статистичної інформації. Для децентралізованої системи гостро постає проблема наукової обґрунтованості методів статистичного спостереження та практичного застосування їх. Утім, незважаючи на специфіку згаданих систем спостереження, кожна з них організується у три основні етапи: підготовка обстеження, його проведення та нагромадження зібраних даних.

Підготовка статистичного спостереження — найвідповідальніший етап, оскільки тут постають і вирішуються головні *методологічні питання*: що і як вивчатиметься, на які запитання мають бути отримані відповіді. На цьому етапі вирішуються *організаційні питання*: хто, де, коли проводить спостереження і які людські, матеріальні та фінансові ресурси для цього необхідно мати. Тобто на першому етапі складається докладний план статистичного спостереження, що містить програмно-методологічні й організаційні питання (рис. 1.2).

На другому етапі збирають дані. Процес копіткий, потребує чіткої взаємодії, скоординованості й оперативності всіх виконав-

чих служб. Чим коротший термін збирання, тим своєчаснішою та дешевішою буде інформація. Але не можна, щоб часові обмеження погіршували якість інформаційного продукту. Від рівня дотримання методологічних та організаційних принципів збирання даних залежать їхня якість, точність, повнота і достеменність. Приміром, якщо неоднозначно визначено статус активних суб'єктів господарювання, це принаймні вплине на те, в який спосіб збиратиметься статистична інформація щодо результатів його господарської діяльності.



Рис. 1.2. Складові плану статистичного спостереження

Третій етап передбачає контроль та накопичення даних спостереження, а також їх збереження. На цьому етапі відпрацьовується система оперативного доступу та пошуку необхідних даних. Це дуже важливо за умов централізованого спостереження, коли нагромаджуються величезні обсяги корисної й дорогої інформації, а знайти її й використати дуже важко, а іноді й неможливо.

Програмно-методологічні питання статистичного спостереження. Помилки спостереження

Розроблення *програмно-методологічних питань* плану спостереження полягає в науково-практичному обґрунтуванні та визначенні суті явища, умов його формування та прояву. Крім того, добирається система ознак, що характеризують явище, враховується можливість їх кількісного опрацювання та перевірки на точність. Комплекс програмно-методологічних питань можна подати в послідовності їх виникнення та розв'язування (рис. 1.3).

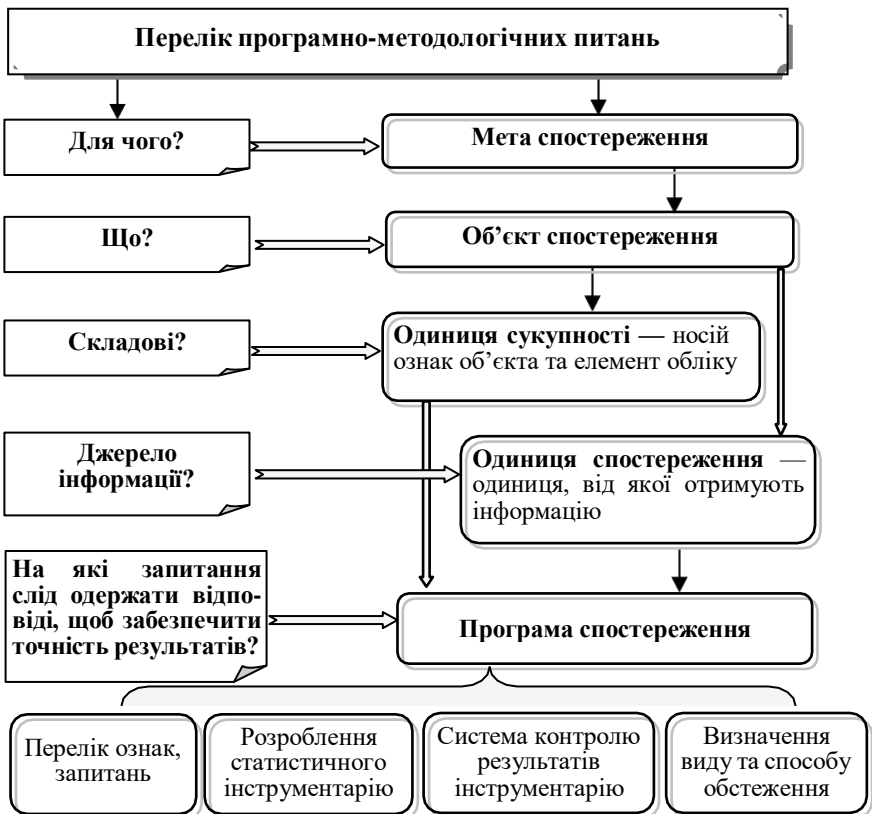


Рис. 1.3. Програмно-методологічні питання плану спостереження

Мета спостереження — отримання статистичних даних, які є підставою для узагальненої характеристики стану та розвитку явища або процесу з визначенням відповідної закономірності. Кінцевою метою спостереження є підготовка управлінських рішень та вжиття відповідних заходів. Так, країни ЄС регулярно провадять обстеження у сфері бізнесу. Метою окремих обстежень є визначення інвестиційної поведінки суб'єктів бізнесу та прогнозування її впливу на рівень виробничих потужностей і зайнятість населення країни. Іншим прикладом може бути обстеження суб'єктів малого і середнього бізнесу в Україні. Мета — ретельне визначення стану виробничо-фінансової діяльності підприємств малого і середнього бізнесу заради сприяння подальшому розвитку їх в Україні.

Мета обстеження визначає його об'єкт. *Об'єкт спостереження* — це сукупність явищ, що підлягають обстеженню. Чітке визначення суті та меж об'єкта дозволяє запобігти різному тлумаченню результатів обстеження. Для цього застосовують набір кількісних обмежувальних ознак, так звані *цензи*. Так, повертаючись до попереднього прикладу, слід обґрунтувати, яка сукупність суб'єктів малого бізнесу буде досліджуватися. Це може бути сукупність усіх зареєстрованих підприємств малого бізнесу або лише діючих. Інша проблема — який суб'єкт бізнесу слід вважати малим? В Україні малим вважається підприємство з чисельністю зайнятих не більше як 50 осіб та обсягом валового доходу від реалізації продукції (робіт, послуг), що не перевищує суми, еквівалентної п'ятистам тисячам євро за середньорічним курсом Національного банку України щодо гривні. Звідси цензом буде максимально припустима кількість зайнятих та обсяг валового доходу.

Об'єкт спостереження як сукупність складається з окремих елементів — одиниць сукупності. *Одиниця сукупності* — це первинний елемент об'єкта, що є носієм ознак, які підлягають реєстрації. Під час перепису населення одиницею сукупності є кожна людина, при реєстрації проданих на фондовій біржі — цінний папір (кожна продана акція, облігація тощо). Проте не кожна одиниця сукупності може надати про себе інформацію. Тому в процесі обстеження виокремлюють одиницю спостереження. *Одиниця спостереження* — це первинна одиниця, від якої одержують інформацію. Так, під час перепису населення одиницею спостереження є домогосподарство, а також кожен його член. При реєстрації проданих на фондовій біржі цінних паперів оди-

ницею спостереження є біржа. Отже, одиниці сукупності та одиниці спостереження можуть збігатися (як, зокрема, під час перепису). І відповідних прикладів існує чимало: в обстеженнях умов життя домогосподарств такими одиницями є домашні господарства, при вивченні громадської думки населення — кожен респондент — людина, яка висловлює свою думку.

Для проведення державного статистичного спостереження економічної діяльності суб'єктів господарювання застосовується також категорія *звітної одиниці*, яка визначається за спеціальними реквізитами й характеризується набором статистичних показників. *Статистичний реквізит* — сполучення цифр і літер, що свідчить про належність одиниці до певної однорідної групи зі спільними властивостями. Літерна інформація може бути подана у вигляді кодових позначень (наприклад, найменування підприємств, видів продукції, видів послуг та їхні коди). Статистичними показниками економічного стану і господарської діяльності виступатимуть: кількість працівників, обсяг реалізованої продукції (товарів, послуг), витрати на виробництво.

Після визначення носіїв ознак і джерел інформації складається програма спостереження. *Програма спостереження* — це перелік запитань, на які слід дістати відповіді в результаті спостереження. Зміст та кількість запитань формуються згідно з метою спостереження і реальними можливостями його проведення (грошовими, трудовими витратами й терміном реєстрації). Складається такий перелік запитань, який за мінімуму залучених даних дає максимум інформації. Цього можна досягти завдяки системному підходу. З одного боку, кожне запитання повинно мати самостійне призначення, а з іншого — кожне запитання в комбінації з іншим має дати додаткову суміжну та побічну інформацію. Складання блоків взаємозалежних узгоджених запитань дає змогу дослідити внутрішні зв'язки об'єкта дослідження. Подальшою деталізацією обраних запитань формується набір ознак, що характеризують зазначений об'єкт обстеження та його одиниці.

Головними вимогами щодо ознак є простота формулювань, чітке тлумачення і визначеність. Так, у переписному листку запитання про шлюбний стан населення потрібно визначити точніше, а саме: замість «перебування у шлюбі» треба «перебуває в зареєстрованому шлюбі». Тоді не буде значних розбіжностей між даними поточного обліку і фактичними даними через незбіг числа фактичних та юридичних шлюбів. Оскільки ознаки за формою

вираження можуть бути не лише кількісними, а й атрибутивними, то, добираючи шкалу вимірювання, перевагу слід віддати не тільки більш інформативним ознакам (номінальній шкалі), а й ознакам із ширшими можливостями статистичного опрацювання (порядковій та метричній шкалі). На підставі сформованого переліку ознак розробляється статистичний інструментарій. *Статистичний інструментарій* являє собою набір статистичних формулярів, а також інструкцій і роз'яснень щодо проведення статистичного спостереження, процедури реєстрації даних.

Статистичний формуляр — це обліковий документ єдиного зразка, що містить адресну характеристику об'єкта спостереження та статистичні дані про нього. Статистичними формулярами є звіти, переписні та опитувальні листки, бланки документів, анкети.

Деякі формуляри містять також коротку інструкцію щодо їх заповнення. До інших інструкція додається як окремий документ, в якому роз'яснюється порядок реєстрації даних, розтлумачується зміст окремих запитань або відповідей. Проте чим даліше формулюються запитання та ознаки, тим менше роз'яснень вони потребують. Так, у процесі складання анкет застосовують запитання закритого, напівзакритого та відкритого типу. До закритого типу належать запитання, до яких додається повний набір відповідей, а респондентові пропонується лише обрати потрібні. Іноді кількість можливих відповідей обмежується (не більше або не менше двох-трьох відповідей). Так, під час соціологічного опитування населення щодо перспектив членства України в Євросоюзі та запровадження безвізового режиму до країн ЄС пропонувалося запитання: «Якщо Ви впродовж року не скористалися безвізовим режимом, то чому?» (дайте не більш як три відповіді).

1. Відсутність потреби і бажання.
2. Недостатність коштів на поїздку.
3. Відсутність біометричного паспорта.
4. Проблеми зі здоров'ям.
5. Вікові обмеження.

Запитання *напівзакритого типу* містять перелік готових відповідей, а також вільний рядок для самовизначення.

У згаданій анкеті про перспективи членства України в ЄС респондентам пропонується запитання: «Що, на Вашу думку, перешкоджає членству України в ЄС?».

1. Низька ефективність реформ.
2. Корумпованість органів влади.

3. Гальмування економічного розвитку.
4. Відсутність гарантій безпеки.
4. Інше _____ (впишіть).

Запитання *відкритого* типу потребують самостійного формулювання відповіді. Такими є, наприклад, запитання: «Якими перевагами членства України в ЄС Ви особисто могли б скористатися?»

При організації масових анкетних спостережень перевагу надають запитанням закритого та напівзакритого типу. Вони дають змогу, по-перше, діставати адекватні відповіді, а, по-друге, — формалізувати відповіді, нагромаджувати їх та кількісно обробляти. Причому більш універсальними є напівзакриті запитання, які передбачають не лише чіткі відповіді, а й пропонують респондентові вільно висловити власну думку.

Статистичний інструментарій забезпечує не лише вхідну частину інформаційної бази спостереження, а й вихідну. Тобто при визначенні ознак та складанні блоків запитань одночасно готуються макети вихідних таблиць, в яких відсутня цифрова інформація. Макети таблиць одразу дозволяють визначити, наскільки кожне запитання (ознака) узгоджується з іншими, передбачити, як «працюватиме» кожне запитання (ознака), виключити малоінформативні запитання, обґрунтувати методику подальшого статистичного опрацювання решти запитань.

Програмою спостереження передбачається також визначення *виду та способу реєстрації даних*. Здебільшого вид і спосіб спостереження залежать від його мети, суті об'єкта спостереження, обсягу та рівня точності очікуваних результатів. Важливими обставинами є також фінансове і трудове забезпечення обстеження, фактор часу. Неодмінною умовою програм спостереження є *наступність* їх змісту, одиниць спостереження, методик обчислення. Це потрібно для порівняння результатів спостереження в часі та просторі. Наприклад, програми переписів населення останніх десятиліть мають однаковий зміст за головними соціально-демографічними показниками. У програмі нового перепису населення України змінено одиницю спостереження, що зробить результати перепису населення порівнянними із міжнародними показниками.

Важливим питанням підготовки статистичного спостереження є *забезпечення точності даних реєстрації*. Точність результатів досягається, з одного боку, відповідною *системою контролю*, а з іншого, — ретельно відпрацьованим механізмом збирання даних

і практичним досвідом у цій роботі. Такий досвід формується під час пробних, так званих *пілотних обстежень* — невеликих за обсягом, що мають на меті випробувати, уточнити програму спостереження, підвищити якість розробки організаційних питань або одержати попередні відомості про маловідоме чи недосліджене явище, процес.

На етапі збирання даних, а нерідко і при підготовці обстеження виникають різного роду помилки. Деяких помилок можна уникнути дотриманням наукових принципів збирання і формування масивів даних. Трапляються випадки, коли про існування окремих помилок не здогадується не лише користувач інформацією, а й сам дослідник. Тому, аби убезпечитися від таких похибок, варто знати причини та природу їхнього походження. Залежно від причини виникнення виокремлюють помилки:

репрезентативності — виникають під час вибіркового спостереження через несущільність реєстрації даних і порушення принципів випадковості відбору;

реєстрації — виникають при будь-якому спостереженні внаслідок неправильного встановлення фактів або неправильного їх запису.

Залежно від природи виникнення *помилки реєстрації* можуть бути випадковими або систематичними.

Випадкові помилки виникають внаслідок збігу випадкових обставин, а саме через неуважність реєстратора або незосередженість респондента. Вони викривлюють дані спостереження в той чи інший бік, проте завдяки величезній кількості випадків їхня дія врівноважується і вони істотно не впливають на результати.

Більш небезпечними є *систематичні помилки*, які виникають в результаті постійних спотворень в одному напрямі. Такі помилки істотно зміщують результати спостереження в один бік (збільшення або зменшення). Так, існує тенденція округлення розмірів доходів і витрат до цілих чисел (доходів у бік зменшення, а витрат — у бік збільшення).

Систематичні помилки бувають *ненавмисними* та *навмисними*. Ненавмисні помилки виникають внаслідок необґрунтованості програми спостереження, некомпетентності реєстраторів, неосвіченості респондентів. Так, через помилки у складанні програми спостереження впливу наслідків аварії на ЧАЕС на стан здоров'я населення результати значно завищили рівень захворюваності, оскільки було обстежено лише населення, яке перебувало в меди-

чних установах. Класичним прикладом помилки, що виникає через неосвіченість респондентів, є так зване жіноче кокетування (зменшення віку), старече кокетування (збільшення віку) та округлення віку до чисел, кратних 5 чи 10, — так звана вікова акумуляція.

Навмисні помилки виникають через свідоме викривлення фактів з певною метою (очорнити або прикрасити дійсність). Прикладами таких помилок є масові факти приховування під час декларування доходів державних службовців, тіньові виплати і розрахунки роботодавців із працівниками, приписки у звітах, приховування реального розміру виторгу від реалізації в комерційній торговельній мережі, пунктах громадського харчування.

Проте навмисні помилки іноді робляться заради самозахисту. Наприклад, з метою уникнення переслідувань та репресій деякі респонденти не визначали або приховували своє віросповідання, етнічну належність, стосунки із родичами, вид діяльності тощо. Із розвитком демократизації суспільства підґрунтя для таких викривлень поступово зникає.

Контроль даних спостереження водночас є методологічним і організаційним питанням. Якщо його розглядати як засіб запобігання помилкам, то питання наближається до методологічного. Якщо контроль — це виявлення й виправлення помилок, то це радше організаційний захід.

Контроль — це насамперед перевірка даних обстежень на їх повноту й достовірність. Аналіз повноти даних здійснюється, як правило, візуально, при перевірці наявності даних за всіма одиницями та позиціями, або порівнянням інформації про об'єкт за різними джерелами даних. Наприклад, неповне охоплення звітністю суб'єктів господарювання можна виявити зіставленням надісланих матеріалів з адміністративними даними, або повнота поточного обліку народжень та смертей контролюються матеріалами переписів населення чи спеціально організованими вибірковими обстеженнями.

Дані на достовірність перевіряють засобами логічного та арифметичного контролю. *Логічний контроль* — це перевірка сумісності даних порівнянням взаємозалежних ознак або виявлення *аутлаєрів*, тобто значення ознаки, яке виходить за ймовірні межі. Прикладом перевірки сумісності даних може бути контроль відповідей респондентів, отриманих у перебігу опитування: порівняння віку респондента з його сімейним станом, виду діяльності із джерелом засобів існування, рівня матеріального становища господарства зі структурою бюджетних витрат тощо. Простим прикладом

виявлення ауглаєрів є визначення віку випускників середніх шкіл при обстеженні якості їхньої освітньої підготовки, оскільки він не може виходити за межі 16–18 років. Зрозуміло, що логічний контроль встановлює лише наявність помилки, а не її розмір. Проте іноді вдається визначити наближені межі помилки, порівнюючи дані спостереження з аналогічними даними інших спостережень чи дані в динаміці. Так, можна визначити помилку у відомостях про котирування цін на біржі, якщо знати курс цін і межі їх коливань; або в розмірі кредитних процентних ставок, коли відома облікова кредитна ставка Нацбанку. Практично неможливо виявити помилку *інлаєр*, значення якої перебуває в логічно допустимих межах. У такому разі застосовують спеціальні статистичні методи моделювання. Порівнянням фактичних даних із модельованими знаходять так звані «викиди», тобто нетипові значення ознаки.

Редагування даних спостереження може бути суцільним або вибіркоvim. Певні переваги має вибіркова перевірка суперечливих даних. При цьому беруть до уваги те, наскільки істотно вплине на загальну картину вибірково одиниця сукупності чи певна її ознака.

Встановити розмір помилки та виправити її можна в процесі *арифметичного контролю*, тобто перевірки прямим чи побічним перерахунком зареєстрованих даних. Наприклад, розмір сукупного доходу домогосподарства можна обчислити з певною точністю, підсумувавши величину основних статей витрат. Розмір фонду заробітної плати на підприємстві чи в установі можна визначити, якщо відомо кількість зайнятих та їхню середньомісячну заробітну плату. Втім, коли помилки виходять за межі логічного й арифметичного контролю, встановити достовірність даних неможливо.

У процесі аналізу та контролю даних спостереження часто виникає необхідність у відновленні загублених чи в заміні відсутніх даних. При цьому використовують спеціальні статистичні прийоми та методи. Одні прийоми застосовують для «компенсації» відсутніх даних для окремих одиниць сукупності, інші — для відновлення втрачених чи відсутніх даних.

У випадку виправлення та заповнення пропусків, тобто «імпутації» даних у сукупності, аналізується їхній розподіл, а саме характер розсіювання значень ознак. Нерідко пропущені дані замінюють на середні значення ознаки для обстеженої сукупності в цілому. Якщо ресурси дозволяють проведення додаткового обстеження, то пропущені дані замінюють, спираючись на результати дообстеження.

Відновлення цілісності (неперервності) часових рядів здійснюють спеціальними прийомами статистичного моделювання, зокрема, для заповнення втрачених раніше даних всередині часового ряду використовують інтерполяцію, а даних, що виходять за межі часового ряду, — екстраполяцію на підставі трендової моделі.

Отже, статистичне спостереження є надзвичайно відповідальним етапом дослідження, оскільки від якості зібраних даних залежать результати подальшого їхнього аналізу.

Організаційні питання статистичного спостереження

Другою складовою плану спостереження є низка організаційних питань. *Організаційні питання* визначають місце, час проведення обстеження, органи та персонал, залучений до обстеження, його матеріально-технічне забезпечення (рис. 1.4).



Рис. 1.4. Організаційні питання плану спостереження

Організаційні питання пов'язані з програмно-методологічними і залежать від мети та умов обстеження. Насамперед з'ясовують, на який орган покладено відповідальність за проведення обстеження, його підготовку. Залежно від масштабу об'єкта спостереження, а також зацікавленості щодо його результатів, можна виокремити певні групи.

1. Центральні органи державної статистики, а саме Державна служба статистики України та її регіональні відділення — державні обстеження на макрорівні. До таких обстежень належать переписи (населення, у сільському господарстві, інноваційних технологій тощо), обстеження соціально-демографічного й економічного характеру (державні структурні обстеження суб'єктів господарювання про їхній економічний стан і господарську діяльність, структурні обстеження кредитних установ, обстеження підприємств з питань оплати праці, обстеження економічної активності населення та умов життя домогосподарств тощо).

2. Статистичні відділи міністерств і відомств — державні обстеження локального за тематикою характеру. Наприклад, обстеження, що їх проводять Державна фіскальна та митна служба, Державна міграційна служба тощо.

3. Спеціальні інститути, агентства, міжнародні організації — обстеження, що ґрунтуються на вивченні громадської думки або мотивації, поведінки та оцінок окремих суб'єктів суспільно-економічного життя. До таких органів належать: Департамент статистики ООН, Київський міжнародний інститут соціології, Міжнародна організація праці (проблеми зайнятості, безробіття, соціального захисту), Міжнародний центр дослідження проблем тенденцій діяльності економічного характеру, Генеральне агентство у справах економіки та фінансів при комітеті ЄС (обстеження у сфері бізнесу, інвестицій, споживання).

4. Аналітичні відділи окремих економічних структур (підприємств, організацій, фірм, банків, бірж, страхових компаній тощо) — обстеження на мікрорівні, що мають маркетингове або контролювальне спрямування. Наприклад, маркетингові обстеження туристичних фірм з метою з'ясування сегмента ринку споживачів або обстеження страхових компаній з метою аналізу потреб фактичних і потенційних клієнтів щодо надання страхових послуг.

Кожен орган спостереження має забезпечити себе кваліфікованими кадрами як на етапі підготовки обстеження, та його

проведення, так і під час опрацювання результатів. Тому в статистичному спостереженні беруть участь не лише статистики (і частіше не статистики), а й представники різних професій. Наприклад, до проведення перепису населення залучаються спеціально підготовлені обліковці-реєстратори, статистики-науковці, організатори та аналітики, а також фахівці у сфері ІТ, технічний персонал.

Наступним питанням є обґрунтування *місця обстеження* — пункту, в якому перебуває одиниця спостереження і де здійснюється реєстрація даних. Наприклад, під час телефонного опитування думки користувачів житлово-комунальних послуг це є місце проживання респондентів. Місцем реєстрації демографічних подій: смертей, народжень, шлюбів та розлучень буде районне відділення реєстрацій актів цивільного стану (РАЦС).

Щоб спостереження забезпечило достовірні та своєчасні дані, необхідно вирішити питання часу та періоду спостереження (див. рис. 1.5).



Рис. 1.5. Час статистичного спостереження

Час спостереження (об'єктивний час) — це час, до якого належать дані спостереження. Якщо об'єктом спостереження є процес, то обирають інтервал, за який нагромаджуються дані. Якщо об'єктом спостереження є певний стан, то обирають *критичний момент* — момент, станом на який здійснюється реєстрація даних. Наприклад, реєстрація числа угод, укладених на універсальній біржі, може відбуватися як за деякий період (упродовж місяця), так і за станом на певні моменти — дні біржових торгів, що відбуваються, як правило, раз на тиждень. Тут первинне спостереження реєструє угоди за моментами, а вторинне — за період.

Критичним моментом користуються в разі перепису населення, коли необхідно точно зареєструвати його стан в один «момент».

Наприклад, критичним моментом перепису населення є час: 24.00 з 20 на 21 січня. Зрозуміло, що провести перепис усього населення країни одночасно в одну хвилину неможливо, тому крім часу спостереження встановлюється *період спостереження* (суб'єктивний час) — час, упродовж якого реєструються дані. Так, період перепису населення становить 10 днів (з 21.01 до включно). Якщо під час перепису обліковець прийшов у домогосподарство 23-го січня, то всі дані про його членів він має реєструвати станом на початок перепису (21.01). Якщо в домогосподарстві на день реєстрації, але після початку перепису народилася дитина, переписом вона не враховуватиметься.

Час спостереження обирають у найсприятливіший або нейтральний для об'єкта спостереження період. Наприклад, перепис плодово-ягідних насаджень здійснюють у період їхнього цвітіння або в період плодоношення. Переписи населення призначаються на період найменшої міграційної активності населення. Для України — це зима, післясвятковий період (кінець січня, початок лютого). Опитування громадської думки провадять у період стабільної, не загостреної соціально-економічної ситуації.

Проведення будь-якого статистичного спостереження потребує відповідного матеріально-технічного забезпечення: обчислювальної та множувальної техніки, транспортних засобів, засобів друкування, статистичного інструментарію та рекламних носіїв. Останнє важливо при підготовці макроспостережень, які потребують роз'яснювань і повідомлень для широкого загалу.

Вправи та запитання для самоконтролю

1. Визначте вид статистичного спостереження за ступенем реєстрації даних:

- а) звіт про діяльність місцевих органів самоврядування області;
 - б) щомісячна звітність про результати фінансової діяльності суб'єктів малого та середнього бізнесу;
 - в) перепис сільськогосподарських тварин в усіх категоріях господарств.
2. Визначте рівень узагальнення статистичних даних:
- а) вартість експортно-імпортних операцій країни;
 - б) розмір прибуткового податку з громадян одного з районів міста;
 - в) обсяг надходжень до Пенсійного фонду країни.

3. Визначте, які вимоги щодо якості статистичних даних було порушено при проведенні спостережень:

- а) аналітичний звіт про масштаби минулорічного спалаху захворювання дітей на кір у шкільних закладах району;
- б) реєстрація незайнятих осіб, які впродовж календарного року зверталися до органів державної служби зайнятості;
- в) звіт про результати зимової сесії студентів-магістрів університету всіх форм навчання (денної, заочної та дистанційної).

4. Зазначте мету та об'єкт спостереження:

- а) перевірка суб'єктів господарювання з питань дотримання податкової дисципліни;
- б) звітність будівельних підприємств про обсяги освоєних інвестицій в основний капітал;
- в) обстеження АЗС регіону стосовно роздрібного продажу світлих нафтопродуктів і газу.

5. Зазначте мету, об'єкт та одиницю спостереження:

- а) моніторинг продажу на аукціоні держоблігацій внутрішнього займу;
- б) складається картотека органів страхування безробітних;
- в) опитування населення щодо самооцінки стану здоров'я в рамках обстеження умов життя домогосподарств.

6. Проводиться опитування контрактних абонентів компанії мобільного зв'язку щодо якості наданих послуг:

- а) чи збігаються в цьому спостереженні одиниця сукупності та одиниця спостереження;
- б) яким має бути це спостереження за мірою охоплення одиниць сукупності та часом реєстрації фактів?

7. Визначте об'єкт, одиницю спостереження та одиницю сукупності обстежень:

- а) вивчення інвестиційної привабливості об'єктів готельного бізнесу в регіоні;

б) облік громадян, які отримали статус внутрішньо переміщених осіб.

8. Визначте, що є цензом при:

а) реєстрації осіб, які звернулися по допомогу до центрів соціальних служб для молоді;

б) веденні реєстру виборців України для громадян правоздатного віку;

в) опитуванні жінок дітородного віку, які перебувають у зареєстрованому шлюбі.

9. Складіть перелік запитань, що входять до програми спостережень:

а) вибіркового обстеження домогосподарств щодо середньомісячного розміру та структури споживання продуктів харчування;

б) опитування учасників тренінгу щодо спрямованості навчального процесу на практичну діяльність;

в) звіту туристичних агенцій про кількість наданих послуг за напрямками подорожей, метою й тривалістю поїздок.

10. Складіть проект статистичного формуляру для обстежень:

а) поточного обліку клієнтів страхової компанії, що має з'ясувати їхній статус (юридична чи фізична особа), матеріальне становище, вид і термін страхування;

б) опитування щодо задоволеності студентів економічних спеціальностей якістю навчання у закладах вищої освіти, з урахуванням статі, курсу, форми навчання, освітнього рівня (бакалавр, магістр), спеціалізації, а також потреб і побажань респондентів;

в) поточного обліку емігрантів з метою з'ясування їхнього соціально-демографічного складу, мети виїзду та країни прибуття.

11. Визначте місце та органи спостереження для організації спостережень:

а) облік домашньої птиці, яка утримується в господарствах населення;

б) моніторинг за рівнем води в річках для запобігання надзвичайним ситуаціям під час весняних паводків;

в) реєстрація державними митними службами обсягу перевезеної громадянами валюти.

12. Визначте органи спостереження та персонал для організації спостережень:

а) облік обсягів купівлі-продажу іноземної валюти в обмінних пунктах мережі комерційних банків;

б) реєстрація співробітниками МВС та СБУ зброї, вилученої в організованих злочинних угрупованнях.

13. Визначте місце, об'єктивний та суб'єктивний час спостереження, а також критичний момент:

а) опитування учасників щорічного книжкового ярмарку з 15.10 до 20.10;

б) контрольний перепис, що здійснюється за станом на 24.00 з 20 до 21 грудня впродовж 10 днів;

в) реєстрація доходів фізичних осіб за підсумками минулого року, що здійснюється впродовж трьох місяців, але не пізніше 30.04 поточного календарного року.

14. Визначте об'єктивний та суб'єктивний час спостереження, а також критичний момент:

а) реєстрація новонароджених, яка здійснюється не пізніше трьох місяців від дня народження;

б) звіт про підсумки сівби зернових культур подається раз на рік не пізніше 5 днів після закінчення сівби;

в) дані обліку дебіторської заборгованості комерційних банків на початок року мають бути подані впродовж 15 днів від початку наступного календарного року.

15. Проведіть логічний контроль правильності заповнення анкети учасника наукової конференції:

Стать	Жіноча
Освіта	неповна вища
Науковий ступінь	доктор наук
Посада	завідувач кафедри
Місце роботи	науково-дослідницький інститут
Мови, якими вільно володієте	англійська, російська
Мова доповіді	українська
Дата приїзду	27.04.2019
Дата від'їзду	31.04.2019
Чи плануєте виступити з доповіддю	ні

16. Проведіть логічний контроль правильності заповнення анкети трудового мігранта України:

Місце народження	м. Рівне
Вік, років	67
Стать	чоловіча
Громадянство	без громадянства
Сімейний стан	одружений

Рівень освіти	вища
Характер візи	одноразова
Правовий статус перебування за кордоном	без офіційного статусу
Тривалість перебування за кордоном	12 місяців
Статус зайнятості	найманий працівник підприємства
Форма трудового договору	письмова

17. Проведіть арифметичний контроль результатів державного статистичного спостереження щодо постачання і використання енергії від відновлюваних джерел:

Вид джерела	Обсяг постачання у році, тис. т н.е.		Структура постачання у році, у % до підсумку		Темп зростання обсягу постачання, %
	минулому	поточному	минулому	поточному	
Гідроенергетика	662	770	18,3	19,5	116,7
Біопаливо та відходи	2830	3000	78,3	76,5	97,7
Вітрова та сонячна енергія	120	150	3,4	3,8	121,0
Загалом	3614	3920	100,0	100,0	108,5

18. Проведіть арифметичний контроль аналітичного звіту сільськогосподарських підприємств про витрати кормів на відгодівлю великої рогатої худоби за два звітні роки:

Рік	Витрачено кормів, тис. т умовних кормо-одиниць		Поголів'я великої рогатої худоби, тис. голів	Витрати кормів на 1 голову великої рогатої худоби, ц
	всього	у тому числі концентрованих		
Минулий	15200	42000	4900	31
Звітний	14500	4550	4100	32

19. Зазначте, яких видів помилок (реєстрації, репрезентативності) припущено при проведенні спостережень:

а) добового обсягу перевезень пасажирів міським автотранспортом тільки комунальної форми власності;

б) добового обсягу перевезень пасажирів міським автотранспортом приватної форми власності в години найменшого навантаження;

в) лабораторних обстежень проб води з артезіанських свердловин із використанням неякісних реактивів.

20. Зазначте, яких видів помилок (випадкові, систематичні) припущено при проведенні спостережень:

а) під час опитування відвідувачок салонів краси частина жінок заокруглювала свій вік у бік зменшення до цифр, кратних п'яти;

б) під час актуалізації даних реєстру виборців помилково було включено осіб, які не досягли 18-тирічного віку;

в) під час обстежень умов життя домогосподарств, у сільській місцевості не було враховано у статті доходів виплати субсидій на придбання скрапленого газу.

Рекомендована література

1. *Гладун О.М.* Вибіркові обстеження населення: методологія, методика, практика : Монографія. — Ніжин: ТОВ «Видавництво «Аспект — Поліграф», 2008. — 344 с.

2. *Гончар О.В.* Забезпечення якості статистичної інформації: методологія та організація : Монографія. — К.: ВПД «Формат», 2011. — 372 с.

3. *Джессен Р.* Методы статистических обследований. — М. : Финансы и статистика, 1985. — 478 с.

4. *Єріна А.М., Пальян З.О.* Статистика : Підручник. — К.: КНЕУ, 2010. — 351 с.

5. *Кузнєцова Н.В.* Виявлення та оброблення невизначеностей у формі неповних даних методами інтелектуального аналізу // Системні дослідження та інформаційні технології. — 2016. — № 2. — С. 104–115. URL: dspace.nbuv.gov.ua/handle/123456789/1340

Розділ 2

ФОРМИ, ВИДИ ТА СПОСОБИ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ

Організаційні форми статистичного спостереження:
сутність, сучасні завдання і перспективи розвитку
Види статистичних спостережень за рівнем охоплення та
асом реєстрації даних
Способи збирання та реєстрації статистичних даних
Сучасна практика застосування комбінованих видів і
пособів статистичного спостереження

Після вивчення цього розділу студент має

- *знати*: сучасні організаційні форми статистичного спостереження, їхні властивості і складові; можливі напрями гармонізації даних різних форм обстеження і об'єднання їх в єдину інформаційну систему нагромадження, опрацювання, аналізу і доступу;
- *уміти*: відповідно до завдань дослідження обирати найефективніші види і способи статистичного спостереження; застосовувати багаторівневі комбіновані форми, користуючись різними джерелами статистичних даних.

Організаційні форми статистичного спостереження: сутність, сучасні завдання і перспективи розвитку

Розмаїття сфер спостереження та багатогранність його аспектів зумовлюють потребу застосовувати різні організаційні форми, види і способи збирання даних. На сучасному етапі вирізняють три організаційні форми спостереження: звітність, спеціально організовані спостереження та реєстри. Їхні властивості і складові ілюструє рис. 2.1.

Звітність — це форма спостереження, за якої кожен суб'єкт діяльності регулярно подає свої дані в державні органи статистики та відомства у вигляді документів (звітів) спеціально затвердженої форми. Звітність характеризується такими властивостями, як обов'язковість, систематичність, достовірність.

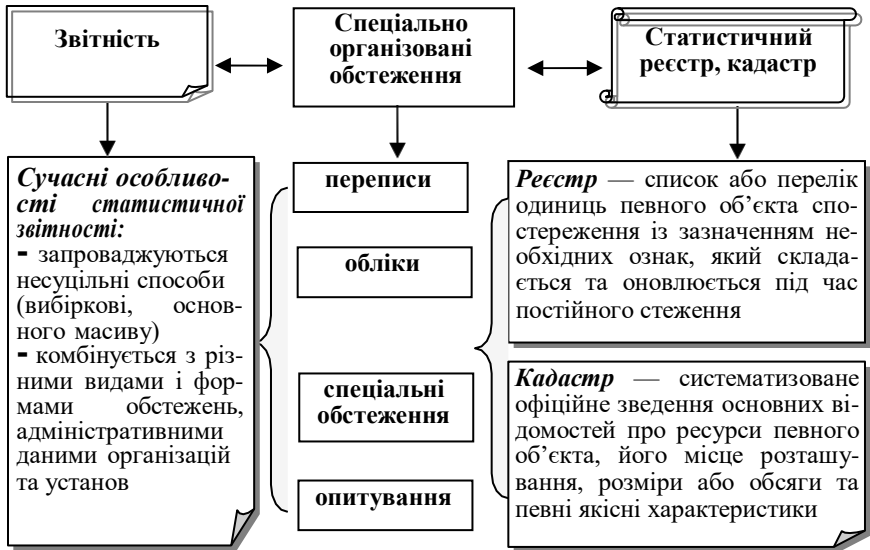


Рис. 2.1. Організаційні форми статистичного спостереження

Обов'язковість — подання звітів обов'язкове для всіх зареєстрованих суб'єктів діяльності із додержанням уніфікованої форми, затвердженого переліку показників, із зазначенням реквізитів підзвітного суб'єкта (назви, адреси, прізвища та підпису відповідальної особи, дати складання звіту).

Систематичність передбачає регулярне, своєчасне складання та подання звітності в затверджені терміни.

Достовірність — дані, наведені у звітності, мають відповідати дійсності й виключати будь-які викривлення (приховування та приписки). За достовірність поданих даних суб'єкти діяльності несуть юридичну відповідальність. Це стає на заваді порушенням через неподання даних чи подання з перекрученням фактів.

Звітність складається на підставі первинних даних оперативного та бухгалтерського обліку. Залежно від рівня затвердження та призначення звітність поділяється на зовнішню та внутрішню. Зовнішня затверджується та збирається органами державної статистики, міністерствами та відомствами, внутрішня розроб-

ляється самим суб'єктом діяльності для власних оперативних, управлінських та аналітичних потреб. Наприклад, аналіз ринкової ситуації, визначення власних ресурсів, прогнозування діяльності ґрунтується на даних внутрішньої звітності багатоструктурних суб'єктів діяльності, акціонерних товариств, концернів, асоціацій тощо.

За частотою подання звітність поділяється на періодичну та річну. Періодична звітність (місячна, квартальна, піврічна) охоплює показники поточної діяльності суб'єктів, річна підбиває головні підсумки фінансової й господарської діяльності суб'єктів за рік. Залежно від терміновості подання звіти можуть передаватися поштою, електронною поштою або через мережу інтернет (електронна звітність).

Останнім часом звітність значно спростилася за формою, скоротилась за обсягом, мірою охоплення, переліком показників. Водночас удосконалюється методика обчислення наявних показників згідно з міжнародними стандартами, розробляється система нових показників, адаптованих до ринкової економіки. Це потребує не лише розроблення системи статистичних показників, а й методів збирання даних, ширшого застосування спеціально організованих спостережень на макро- та мікрорівні.

Спеціально організовані спостереження — це форма спостереження, яка охоплює сфери життя та діяльності, що не вловлюються звітністю. До таких спостережень належать: переписи, обліки, спеціальні обстеження, опитування.

Перепис — це суцільне або рідше вибіркове спостереження масових явищ з метою визначення їх розміру та складу на певну дату. Перепис здійснюється періодично (як правило, з рівним інтервалом) чи одноразово. Так, переписи населення в більшості країн світу провадять кожні 10 років у роки, що закінчуються на «0», «1» чи «9». Особливістю переписів є те, що вони проводяться одночасно по всій території за єдиною для всіх одиниць програмою. До програми перепису населення України включено запитання, що стосуються окремої людини (вік, стать, етнічне походження, рідна мова, шлюбний стан, освіта, громадянство, джерела засобів існування), а також домогосподарства в цілому (склад домогосподарства, зайнятість його членів, житлові умови тощо).

Обліки — суцільні спостереження масових явищ, які ґрунтуються на даних огляду, опитування та документальних записів.

Прикладом можуть бути обліки поголів'я худоби за видами, групами і категоріями господарств, а також обліки земельного фонду за видами угідь, якістю ґрунту, землевласниками та землекористувачами, обліки запасів матеріальних ресурсів на складах або готівки в касах.

Спеціальні обстеження — несуцільне спостереження окремих масових явищ згідно з певною тематикою, що виходить за межі звітності. Вони можуть бути періодичними або одноразовими. Наприклад, обстеження населення з питань упровадження реформи охорони здоров'я, обстеження домогосподарств з питань економічної активності, маркетингові обстеження.

Опитування — це, як правило, несуцільне спостереження думок, мотивів, оцінок, що реєструються зі слів респондентів. Винятком є суцільне опитування всього населення — *референдум* — масове волевиявлення щодо принципових соціально-політичних та економічних питань. Так, 1991 року відбувся загальнодержавний референдум щодо проголошення незалежності України. Інший приклад — референдум з питання приєднання України до Євроатлантичного альянсу та Євросоюзу.

Опитування можуть здійснюватися в різних формах: усній (інтерв'ю), письмовій (анкетування), очній (роздача анкет), заочній (поштової, телефонні). Останнім часом у світовій практиці широко застосовують телефонні опитування. Наприклад, опитування споживачів про їхні наміри щодо придбання певних товарів, користування традиційними і новими послугами, про споживчу мотивацію, прихильність або задоволеність продукцією конкретного виробника чи якістю наданої послуги, про ставлення до реклами тощо.

Відродилася така форма спостереження як *статистичний реєстр* — список або перелік одиниць певного об'єкта спостереження із зазначенням необхідних ознак, який складається та оновлюється під час постійного відстежування. Перші реєстри склалися в Російській імперії упродовж XVIII–XIX століть під назвою «рев'язькі сказки» — поіменний список мешканців країни із зазначенням їхнього віку, статі, сімейного та майнового стану.

У наш час діє велика кількість державних реєстрів, без яких важко уявити єдину державну інформаційну систему. Одним з найперших і основних для спостереження за фінансово-економічною діяльністю всіх суб'єктів господарювання був

створений Єдиний державний реєстр підприємств та організацій України (ЄДРПОУ), що дозволяє здійснювати на державному рівні єдиний облік та ідентифікацію всіх зареєстрованих юридичних осіб за формами власності та організаційно-правовими формами господарювання. Дані реєстру використовуються для аналізу структурних зрушень серед суб'єктів господарювання внаслідок реорганізації, виникнення нових та ліквідації раніше створених юридичних осіб. Цей реєстр дає змогу налагодити єдиний інформаційний простір, в який входять зареєстровані суб'єкти ринку. Паралельно функціонують: Єдиний державний реєстр юридичних осіб, фізичних осіб-підприємців та громадських організацій, Державний реєстр фізичних осіб — платників податків, що містить дані про суми одержаних доходів та утриманих податків.

Кадастри — систематизоване офіційне зведення основних відомостей про ресурси певного об'єкта, його місце розташування, розміри або обсяги та певні якісні характеристики. Прикладом може бути Державний земельний кадастр — комплекс відомостей та документів, що містять дані про природний, господарський і правовий стан земель для забезпечення раціонального використання земельних ресурсів країни. Він, зокрема, містить відомості про кількість земель, їхній картографічний та геодезичний опис, якість і родючість угідь, адміністративно-територіальну належність земель, вартісну оцінку різних категорій земель. Водний кадастр — система зведених гідрологічних даних про водні об'єкти, що використовуються для оцінювання водних ресурсів, забезпеченість ними та використання на певних територіях.

Сучасна стратегія Держстату України щодо проведення статистичних спостережень передбачає чимало новацій, пов'язаних з переходом державної статистики на міжнародні статистичні стандарти. Зокрема, подальшого поширення набуватимуть статистичні реєстри, спеціальні обстеження несудільного характеру, комбіновані організаційні форми (звітність, обліки, опитування тощо). А це, своєю чергою, висуватиме нові завдання — гармонізацію даних різних форм обстеження і об'єднання їх в єдину інформаційну систему: збирання, нагромадження, опрацювання, аналізу і доступу.

Види статистичних спостережень за рівнем охоплення та часом реєстрації даних

При класифікації статистичних спостережень визначають їх вид і спосіб реєстрації даних. Види спостереження розрізняють за двома критеріями: рівнем охоплення одиниць і часом реєстрації даних.

За рівнем охоплення вирізняють суцільні та несуцільні спостереження (рис. 2.2). *Суцільні спостереження* — це обстеження, при яких реєструються всі без винятку одиниці сукупності. До цього виду належать обстеження у формі звітності, розрахованої на певних суб'єктів діяльності, а також більшість переписів. Винятком є перепис населення, який поєднує суцільне та вибіркове обстеження за окремим переліком ознак.

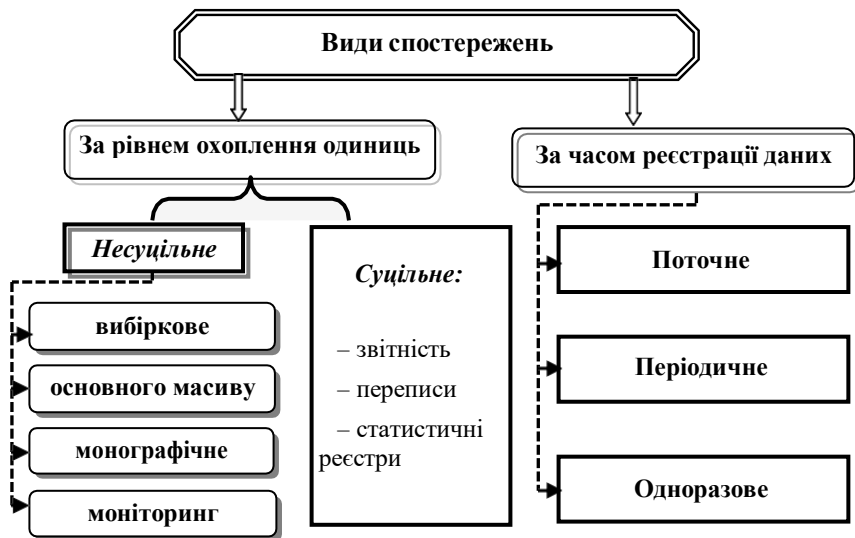


Рис. 2.2. Види статистичних спостережень

Несуцільні спостереження — це обстеження, при яких реєструються не всі одиниці сукупності, а лише їх певна частина. Серед таких спостережень вирізняють: вибіркове, основного масиву, монографічне, анкетне, моніторинг.

Вибіркове спостереження — це обстеження, при якому реєструється певна частина одиниць сукупності, дібрана у випадково-

му порядку. Прикладом можуть бути вибіркові обстеження суб'єктів малого бізнесу, обстеження рівня знань студентів державних і недержавних закладів вищої освіти, обстеження економічної активності населення, доходів та витрат домашніх господарств, їхніх умов життя, а також обстеження якості товарів і продукції, попиту на певні соціальні послуги тощо.

Обстеження основного масиву — це обстеження переважної частини одиниць сукупності, що відіграють визначальну роль в характеристиці об'єкта спостереження. Прикладом може бути обстеження міст з найвищим рівнем забруднення атмосферного повітря або обстеження діяльності групи найвпливовіших фінансових корпорацій.

Монографічне обстеження — це детальне обстеження окремих типових одиниць сукупності з метою досконалого їхнього вивчення. Прикладами можуть бути обстеження діяльності Пенсійного фонду країни, Фонду державного майна України (ФДМУ), електронної системи державних закупівель Prozorro.

Окреме місце посідає *анкетне спостереження* — це обстеження певної частини одиниць сукупності внаслідок неповного повернення від респондентів заповнених реєстраційних формулярів (анкет). Обстеження може відбуватися як вибіркоче з дотриманням принципів випадковості відбору або без дотримання засад теорії імовірності. До останніх належать «автовибірки», коли несущільна сукупність формується сама по собі або з «легкодоступної» частини одиниць. Прикладом є анкетне обстеження мешканців регіону щодо їхнього ставлення до процесу забудови прибудинкових зелених зон чи обстеження студентів щодо забезпечення належної інформаційно-комунікаційної інфраструктури у закладах вищої освіти.

Моніторинг — це комплексне неперервне спостереження за станом певного середовища або об'єкта згідно зі сталим набором його параметрів чи індикаторів. В економічній сфері моніторинг застосовують при реєстрації даних біржових торгів (валютних, товарних, цінних паперів). Здійснюється також моніторинг цін на внутрішньому та зовнішньому ринках, наприклад цін на пальне, зерно тощо.

У соціальній сфері запроваджено моніторинг соціальних настроїв населення, а також моніторинг якості освіти випускників середніх навчальних закладів. За умов загострення демографічної кризи в країні та погіршення стану здоров'я населення створено

Державний моніторинг соціально-гігієнічного благополуччя населення, а також моніторинг за поширенням окремих видів інфекційних хвороб.

Спостереження за часом реєстрації фактів поділяються на поточне, періодичне та одноразове. *Поточне спостереження* — це систематична реєстрація фактів щодо явищ мірою їх виникнення або збирання фактів щодо безперервного процесу. Безперервними є демографічні процеси: народжуваність, смертність, шлюбність і розлучуваність, а також виробничі процеси (випуск продукції) або реалізаційні процеси (збут і реалізація продукції). До числа явищ, які реєструються в певні моменти, належать підсумки біржових торгів, продажу на аукціонах.

Періодичне спостереження проводиться за певні (як правило, рівні) проміжки часу. Прикладом можуть бути переписи населення, виробничих площ, інноваційних технологій тощо, а також обстеження ділової активності суб'єктів господарювання.

Одноразове спостереження проводиться мірою виникнення потреби в дослідженні явища чи процесу. Наприклад, маркетингове дослідження щодо адаптації товару до місцевого ринку, або обстеження думки населення щодо впровадження страхової медицини чи обґрунтованості введення нових тарифів на житлово-комунальні або транспортні послуги.

Способи збирання та реєстрації статистичних даних

Статистичні спостереження здійснюються трьома способами: безпосередній облік фактів, документальний облік, опитування.

Безпосередній облік — це обстеження, за якого обліковець особисто реєструє факти підрахунком, вимірюванням, оцінюванням, оглядом. Так здійснюється реєстрація товарних потоків, що перетинають митні кордони, облік готівкової грошової маси в банках чи оцінювання рівня екологічного забруднення в регіоні тощо. Проте більшість явищ і процесів суспільно-економічного життя не підлягають прямому вимірюванню. У таких випадках застосовують інші два способи.

Документальний облік — це обстеження, за якого реєстрація фактів відбувається за даними, наведеними в документах первинного обліку. У такий спосіб складається статистична звіт-

ність. Таким чином визначаються всі економічні показники макрота і мікрорівнів: обсяги матеріальних, трудових і фінансових ресурсів, розмір доходів та витрат, інвестицій, обсяги експорту й імпорту товарів та послуг тощо.

Опитування може здійснюватися різними шляхами: експедиційним, самореєстрацією, кореспондентським та анкетним. *Експедиційний спосіб* — це реєстрація фактів спеціально підготовленими обліковцями з одночасною перевіркою точності реєстрації. Такий спосіб застосовується під час переписів населення в Україні. В інших країнах обирають дешевший спосіб — самореєстрацію (США, країни Західної Європи).

Самореєстрація — це реєстрація фактів самими респондентами після попереднього інструктажу з боку реєстраторів-обліковців. У США дані перепису населення, одержані самореєстрацією, дають 67 % повернення заповнених переписних листків. Решта даних добирається експедиційним способом. Отже, метод самореєстрації дає добрі результати там, де якісно ведеться роз'яснювальна робота, де респонденти характеризуються високим рівнем громадянської відповідальності.

Кореспондентський спосіб — це реєстрація фактів про явища та процеси на місцях їхнього виникнення спеціально підготовленими особами та надсилання результатів у відповідні інстанції. Цей спосіб широко використовується при дослідженнях ринку товарів та послуг на рівні окремих регіонів, обстеженні процесу просування товарів у специфічних умовах ринку.

Окремі види та способи спостереження можуть використовуватися в комплексі, не виключаючи один одного, залежно від складності доступу до об'єкта спостереження, рівня підготовленості респондентів до певного методу спостереження, сучасних досягнень з методології й організації статистичних спостережень.

Сучасна практика застосування комбінованих видів і способів спостереження

Різні типи економічних систем і притаманні їм форми господарювання зумовлюють не лише певні запити суспільства на статистичну інформацію, а й відповідні способи її отримання. Так, за традиційного типу економіки, основними способами спосте-

реження були обліки і примітивні форми переписів. За командно-адміністративного типу економічної системи панівною формою спостереження була статистична звітність, що здійснювалася у документальний спосіб і охоплювала всі без винятку адміністративно-господарські одиниці. За ринкової, а тим паче змішаної економіки, якій притаманне поєднання ринкового механізму з державними методами регулювання діяльності різних суб'єктів господарювання, застосовують плюралізм форм, видів і способів спостереження (див. рис. 2.3).

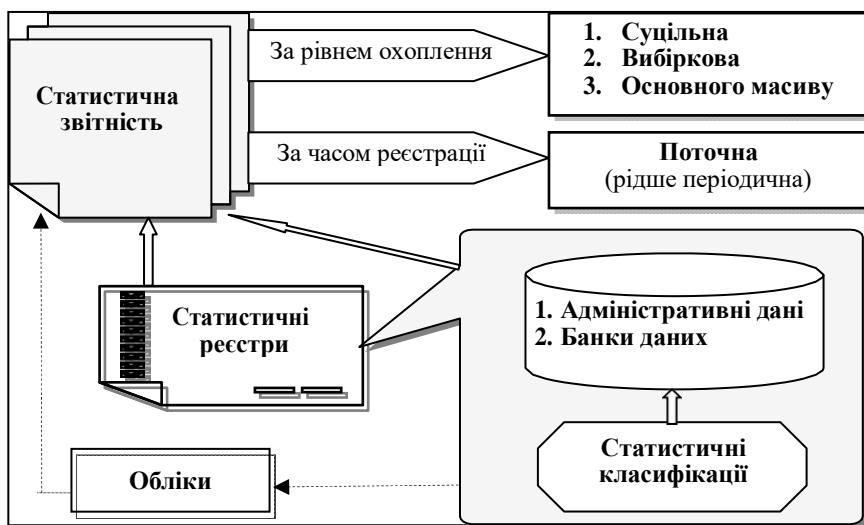


Рис. 2.3. Форми та види статистичних спостережень у комбінації зі звітністю

За сучасних умов гармонізації офіційної статистики України з міжнародними інформаційними стандартами широко запроваджуються багаторівневі комбіновані форми, види і способи статистичних обстежень. Так, докорінної зміни зазнала класична для планової економіки форма спостереження — статистична звітність, яка комбінується з адміністративними даними, розміщеними в різноманітних статистичних реєстрах чи в банках даних державних установ і адміністрацій. Причому за мірою охоплення одиниць звітність тепер зорганізується не лише як суцільна, а й

як обстеження основного масиву чи як вибіркове. Це зменшує звітне навантаження на респондентів, а також заощаджує фінансові й людські ресурси, пов'язані зі збиранням даних.

Прикладом комбінованих обстежень є Державне статистичне структурне спостереження. Його мета — отримання статистичної інформації про економічний стан і господарську діяльність активних суб'єктів господарювання (окрім банківських установ, страхових компаній і недержавних пенсійних фондів). Така інформація використовується для аналізу структури та розвитку суб'єктів господарювання всіх видів економічної діяльності.

За рівнем охоплення одиниць державне структурне спостереження є комбінацією суцільного і вибіркового обстежень, зорганізованих у формі звітності, а також з використанням даних статистичного реєстру та адміністративних даних різних державних установ. Отже, згадане вище комплексне обстеження охоплює різні одиниці спостереження. Передусім, це господарські організації — юридичні особи, комунальні та інші підприємства, створені відповідно до Господарського кодексу України, а також інші юридичні особи, які здійснюють господарську діяльність та зареєстровані в установленому законом порядку.

Другу групу одиниць спостереження становлять фізичні особи-підприємці — громадяни України, іноземці та особи без громадянства, які здійснюють господарську діяльність та зареєстровані відповідно до закону як підприємці. Для обох груп одиниць спостереження складається список звітних одиниць, активних станом на 15 листопада звітного року і розподілених за інституціональними секторами економіки. Список формується на підставі Реєстру статистичних одиниць (РСО) за *ідентифікаційними показниками* — ідентифікаційний код суб'єкта відповідно до реєстрації юридичних осіб в ЄДРПОУ; індивідуальний код місцевої одиниці.

Одиницями спостереження такого обстеження є: підприємство; місцева одиниця; одиниця за видом економічної діяльності (ВЕД); місцева одиниця за ВЕД; фізична особа-підприємець. Кожна з одиниць має чітке визначення за набором ідентифікаційних показників. Так, підприємство трактується як статистична одиниця, що здійснює виробництво товарів та надання послуг, є самостійною у прийнятті рішень щодо використання своїх засобів і найбільш незалежною з огляду на фінансові та виробничі функції. Підприємство здійснює один або кілька видів економічної

діяльності в одному чи кількох місцях. Підприємство складається з однієї або більше місцевих одиниць. Одному підприємству відповідає одна юридична особа. Своєю чергою, місцевою одиницею вважається підприємство або його частина, що розташована в географічно визначеному місці (за поштовою адресою). Місцева одиниця може належати лише одній юридичній особі. Одиниця за видом економічної діяльності (ВЕД) — активне підприємство, що здійснює вид діяльності, на який припадає найбільший внесок у валову додану вартість.

На суцільній основі обстежуються юридичні особи, в яких обсяг реалізованої продукції (товарів, послуг) упродовж календарного року забезпечує 80 % від відповідного загальнодержавного показника. Ця категорія респондентів заповнює річну форму статистичної звітності № 1-підприємництво «Структурне обстеження підприємства». Форма звітності складається з 10 розділів, де зазначаються показники результатів господарської діяльності (обсяг реалізованої продукції, товарів, послуг; вартість продукції власного виробництва для внутрішніх потреб підприємства; витрати на виробництво; вартість товарів і послуг, придбаних для перепродажу; розмір матеріальних витрат капітальних інвестицій у матеріальні активи, витрат на виробництво, залишків запасів; наявність на підприємстві інтегрованої діяльності, коли різні стадії виробництва здійснюються послідовно, тощо). Подаються також дані про кількість працівників та відпрацьований ними робочий час. Окремі розділи стосуються статутного капіталу та його розподілу між засновниками, а також структурних змін, що відбулися на підприємстві у звітному році (збільшення чи зменшення активів підприємства, оренди основних засобів).

Вибіркове обстеження охоплює інші юридичні особи, які не були включені до суцільного, а саме малі підприємства. Вони також подають річну статистичну звітність, але за скорченою формою «Структурне обстеження підприємства», що містить лише 6 розділів. За способом статистичного спостереження в обох випадках застосовується документальний облік на паперових чи електронних носіях (електронна звітність).

З метою контролю показників діяльності юридичних осіб, які охоплюються суцільним та вибіркоvim обстеженням, а також для поширення даних річного вибіркового структурного обстеження малих підприємств на всю генеральну сукупність залучаються додаткові дані з кварталних та річних форм фінансової звітності

(активи, зобов'язання і фінансові результати підприємств), матеріали річної звітності про капітальні інвестиції.

Звітна інформація стосовно юридичних осіб доповнюється адміністративними даними. До *адміністративних даних* належить інформація, яку збирають і нагромаджують інші державні органи, окрім органів державної статистики, а саме органи місцевого самоврядування та юридичні особи, на які покладено виконання відповідних адміністративних функцій. Так, у рамках Державного статистичного структурного спостереження використовуються банки даних, що формуються: Національною комісією з державного регулювання у сфері ринків фінансових послуг, Національною комісією з цінних паперів та фондового ринку (НКЦПФР) з відповідними переліками підприємств — професійних учасників фондового ринку, а також Фондом державного майна України (ФДМУ) — щодо переліку господарських товариств, у яких держава володіє корпоративними правами і які належать до державного сектору економіки.

Джерелом статистичної інформації стосовно фізичних осіб-підприємців є винятково адміністративні дані. А саме використовуються електронні масиви деперсоніфікованих адміністративних даних (банки даних) щодо діяльності фізичних осіб-підприємців, які надає Державна фіскальна служба України (ДФС) в рамках взаємної угоди з Держстатом України. Таким чином отримують статистичні дані про річні показники кількості найманих працівників, доходів, витрат діяльності і сплачених податків (для окремих одиниць відповідно до режиму оподаткування).

Інтеграція всіх згаданих одиниць спостереження та статистичні дані про них в єдину інформаційну систему була б неможливою без застосування національних класифікацій (класифікаторів). У них містяться *класифікаційні показники* — реквізити суб'єкта щодо його належності до певного виду економічної діяльності, території, організаційно-правової форми господарювання тощо.

Для проведення Державного статистичного структурного спостереження використовують такі національні класифікації: КВЕД (класифікація видів економічної діяльності), КОПФГ (класифікатор організаційно-правових форм господарювання), КОАТУУ (класифікатор об'єктів адміністративно-територіального устрою України), КІСЕ (класифікація інституціональних секторів економіки), СКП (статистична класифікація видів продукції).

У багатьох сферах соціально-економічної діяльності застосовуються комбіновані статистичні спостереження, що організуються як спостереження основного масиву в поєднанні з даними суцільних і вибірових обстежень, статистичного обліку та адміністративних даних. Прикладом може слугувати інформаційне забезпечення статистики економічної активності населення, його зайнятості та безробіття. Так, дані про стан виплати заробітної плати працівникам отримують у результаті обстеження основного масиву місцевих одиниць, шляхом подання місячної форми звітності. Інформацію про кількість працівників та відпрацьований ними робочий час збирають в рамках Державного статистичного структурного спостереження (на суцільній та вибірковій основі). Обстеження підприємств з питань статистики праці щодо оплати праці, динаміки кількості найманих працівників та руху кадрів, використання робочого часу, рівня неповної зайнятості, обсягів і структури фонду оплати праці організується як вибірове. Дані подаються у формі звітності (форма 1-ПВ «Звіт з праці» місячна, квартальна).

У процесі вибірових обстежень домогосподарств з питань економічної активності населення щомісяця отримують інформацію про чисельність економічно активного, неактивного населення, зайнятого, у тому числі неформально зайнятого населення, про чисельність безробітних у визначенні Міжнародної організації праці (МОП). Обстеження членів домогосподарств здійснюється заповненням двох анкет: «Анкета домогосподарства» та «Анкета обстеження робочої сили». Паралельно Державною службою зайнятості України здійснюється статистичний облік незайнятого населення і зареєстрованих безробітних, яким надано відповідний статус. Цей орган формує банк адміністративних даних про кількість і соціально-демографічні та економічні характеристики незайнятих і безробітних. Але дані Державної служби зайнятості стосуються лише тієї категорії незайнятого населення, яка звернулася до центру зайнятості, активно шукає і готова приступити до роботи чи навчання за направленням. Суцільна ж реєстрація всього економічно неактивного, незайнятого, неформально зайнятого і безробітного населення, а також трудових мігрантів можлива лише за умови проведення перепису населення. Також переписи є надійним джерелом інформації стосовно виду діяльності населення країни за основною роботою або власної справи, освіти, професії, характеру роботи, виду зайнятості, джерел засобів існування (рис. 2.4).

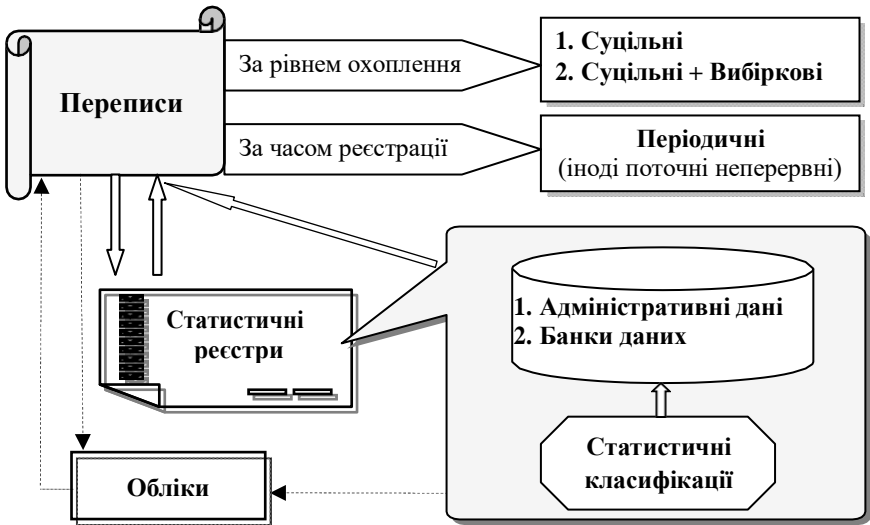


Рис. 2.4. Форми та види статистичних спостережень в комбінації із переписами

Слід зазначити, що переписи нерідко комбінуються з вибірковими обстеженнями. Так, суцільні переписи населення України 1959, 1970, 1979 та 1989 років поєднувалися із вибірковими. Наприклад, програма останнього Всесоюзного перепису населення та житлового фонду 1989 року включала запитання, за якими опитування населення здійснювалося не лише на суцільній основі, а й вибірково (місце роботи, вид занять, належність до суцільної групи, тривалість безперервного проживання тощо). Утім, переписи — дуже дорога форма суцільного обстеження і, як згадувалося раніше, здійснюються періодично через доволі тривалі проміжки часу. Тому в міжпереписний період використовують дані поточного обліку або реєстри чи спеціально зорганізовані вибіркові обстеження. Скажімо, матеріали перепису населення щороку доповнюються й оновлюються даними поточного обліку записів актів цивільного стану (смертей, народжень, шлюбів, розлучень, прибуття та вибуття).

Свою чергою, дані перепису населення і поточного обліку окремих демографічних подій заносять у відповідні реєстри (демографічний реєстр, реєстр виборців, платників податків тощо).

Ця інформаційна база в гармонізованому вигляді використовується для проведення моніторингів: демографічної ситуації і процесу відтворення населення; зміни чисельності і вікового складу контингенту виборців, платників податків, пенсійного, до працездатного і працездатного контингентів тощо. Моніторинг, що спирається на результати несудильних обстежень у формі опитування населення, слугує підставою для прогнозування очікуваних змін у соціальному складі населення, у структурі електорату, платників податків. А це має враховуватися при коригуванні й оновленні наявних інформаційних баз даних.

Іншим прикладом комбінованого спостереження є статистичні обстеження в аграрному секторі економіки. З огляду на розмаїття організаційно-правових форм господарювання в сільському господарстві застосовуються різні види, форми та способи спостереження. Тут статистична інформаційна база формується на підставі оперативної та річної форм звітності щодо вирощування продукції рослинництва, тваринництва, обсягу реалізації сільськогосподарської продукції, механізації виробничого процесу, фінансово-господарській діяльності. Цю звітність подають всі юридичні особи, що займаються виробництвом сільгосппродукції. Тобто респондентами є сільськогосподарські підприємства усіх форм власності, організаційно-правових форм господарювання та підпорядкування, підсобні сільські господарства підприємств та організацій інших галузей економіки, великі селянські (фермерські) господарства, створені на базі колективного сільськогосподарського підприємства.

Паралельно статистика збирає дані про виробництво сільськогосподарської продукції у господарствах населення. Для цього використовують комбінацію таких форм і видів спостереження: погосподарський облік, вибіркові обстеження і тематичні переписи. А саме щорічний погосподарський облік реєструє станом на 1 січня дані про чисельність поголів'я худоби та птиці, а також розмір та структуру посівних площ основних груп сільськогосподарських культур в особистих підсобних господарствах населення сільської місцевості. На регулярній основі проводять вибіркові обстеження сільськогосподарської діяльності населення в сільській місцевості та обстеження умов життя домогосподарств сільської місцевості. Періодично проводяться тематичні сільськогосподарські переписи, зокрема перепис багаторічних насаджень в усіх категоріях господарств.

Вочевидь, що широке запровадження у статистичну практику різноманітних форм і видів спостереження висуває нові завдання — узгодження між собою різних статистичних банків даних, їхнє системне нагромадження та вільний доступ до агрегованої інформації.

Вправи та питання для самоконтролю

1. Визначте форму статистичного спостереження:

№ з/п	Статистичне спостереження	За формою спостереження		
		звітність	спеціально організоване	реєстр
1	Перепис об'єктів незавершеного будівництва			
2	Списки учасників Фонду гарантування вкладів фізичних осіб			
3	Річні звіти про виробничо-фінансову діяльність малого підприємства			
4	Стеження за результатами впровадження економічних реформ у країні			
5	Реєстрація учасників зовнішнього незалежного оцінювання знань			

2. Визначте вид статистичного спостереження за рівнем охоплення одиниць:

№ з/п	Статистичне спостереження	За мірою охоплення одиниць	
		суцільне	несуцільне
1	Облік школярів, які потребують інклюзивної освіти		
2	Опитування студентів закладів вищої освіти щодо їхньої академічної мобільності		
3	Списки виборців регіональних виборчих округів		
4	Онлайн опитування користувачів соціальної мережі «Facebook»		

3. Визначте вид несутільного спостереження:

№ з/п	Статистичне спостереження	Несутільне спостереження		
		вибіркове	основного масиву	монографічне
1	Перевірка результатів діяльності Фонду державного майна			
2	Обстеження якості пального, що відпускається споживачам на автозаправних станціях міста			
3	Реєстрація цін на плодово-ягідну продукцію на організованих продуктових міських ринках			
4	Звіт про результати фінансової діяльності найвпливовіших страхових компаній			

4. Визначте вид несутільного спостереження:

№ з/п	Статистичне спостереження	Несутільне спостереження	
		анкетне	моніторинг
1	Стеження за технологічними параметрами атомних електростанцій в країні		
2	Обстеження думки населення щодо якості роботи центрів надання адміністративних послуг		
3	Стеження за коливанням валютних курсів на біржі		
4	Голосування радіослухачів у соціальних мережах під час формування музичних чартів		

5. Визначте вид статистичного спостереження за часом реєстрації фактів:

№ з/п	Статистичне спостереження	За часом реєстрації фактів		
		поточне	періодичне	одноразове
1	Моніторинг викидів в атмосферу парникових газів			
2	Опитування молодих сімей щодо впливу соціальних виплат на їхні дітородні наміри			
3	Складання списку осіб призовного віку			
4	Реєстрація в Державній службі зайнятості непрацевлаштованих осіб			

6. Визначте спосіб статистичного спостереження:

№ з/п	Статистичне спостереження	Способи спостереження		
		безпосередній облік	документальний облік	опитування
1	Обстеження думки населення щодо ефективності діяльності Об'єднання співвласників багатоквартирних будинків (ОСББ) на Форумі відповідного сайта			
2	Реєстрація органами МВС кількості вогнепальної зброї у населення			
3	Облік незаконного вилову риби у природних водоймах області			
4	Облік наданої цільової допомоги незахищеним верствам населення регіону			

Рекомендована література

1. *Гладун О.М.* Вибіркові обстеження населення: методологія, методика, практика. — Ніжин: ТОВ «Видавництво «Аспект — Поліграф», 2008. — 344 с.

2. *Джессен Р.* Методы статистических обследований. — М.: Финансы и статистика, 1985. — 478 с.

3. *Єріна А.М., Пальян З.О.* Статистика : Підручник. — К.: КНЕУ, 2010. — 351 с.

4. Методика формування вибірових сукупностей для проведення у 2014–2018 роках вибірових обстежень населення (домогосподарств): умов життя домогосподарств, економічної активності населення та сільськогосподарської діяльності населення у сільській місцевості (затверджена наказом Держстату від 08.02.2013 № 37). URL: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2013/37/metod_vyb_suk.zip

5. Методологічні положення з організації державного статистичного спостереження щодо структурних змін в економіці України та її регіонів (затверджені наказом Держстату від 29.07.2016 № 134). http://www.ukrstat.gov.ua/metod_polog/metod_doc/2016/134/134_2016.htm

6. Порядок подання електронної звітності до органів державної статистики (затверджений наказом Держкомстату від 12.01.2011 №3). URL: <https://zakon.rada.gov.ua/laws/show/z0408-11>

Розділ 3

ПЕРЕПИСИ ТА СТАТИСТИЧНІ РЕЄСТРИ: СУТНІСТЬ, ЕВОЛЮЦІЯ, СУЧАСНА ОРГАНІЗАЦІЯ

Переписи як спеціально організована форма статистичного спостереження

Особливості сучасних переписів: міжнародний і вітчизняний досвід

Перепис населення: етапи проведення, програмно-методологічні та організаційні питання

Статистичні реєстри та формування статистичних баз даних

Після вивчення цього розділу студент має

- *знати*: сутність, характерні особливості та сфери використання даних переписів; міжнародний досвід організації переписів; програмно-методологічні та організаційні питання перепису населення; функції державних статистичних реєстрів;
- *уміти*: складати плани статистичного спостереження у формі перепису, зокрема програмно-методологічні питання (мета, об'єкт та одиниця спостереження, одиниця сукупності, перелік запитань, статистичний формуляр, форма контролю результатів), а також вирішувати організаційні питання (місце, час, орган, персонал, вид, форма та спосіб спостереження).

Переписи як спеціально організована форма статистичного спостереження

Як зазначалося вище (2.1), *перепис* — це суцільне чи комбіноване з вибіркового спостереження масових явищ для визначення їхнього розміру та складу для всієї сукупності одиниць станом на певну дату. Зазвичай перепис не обмежується збиранням даних, він також включає комплекс робіт, пов'язаних з опрацюванням, аналізом, опублікуванням і поширенням даних.

Перепис як форма статистичного спостереження має характерні особливості (рис. 3.1).

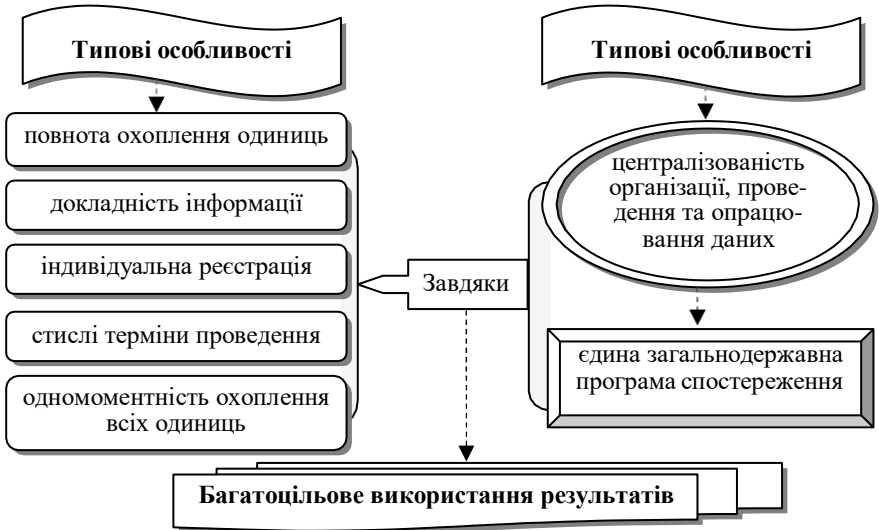


Рис. 3.1. Типові особливості переписів

Повнота охоплення, що власне і є визначальною рисою перепису як суцільного обстеження, означає, що обліковуватися має кожна, без винятку, одиниця спостереження на всіх територіях чи об'єктах, які належить обстежити. На практиці не завжди вдається забезпечити стовідсоткове охоплення. Наприклад, якщо на момент перепису одиниця спостереження відсутня або відмовляється від участі в ньому, мотивуючі наявністю значної частини інформації в інших джерелах, таких як адміністративні реєстри, або конфіденційністю персональних даних чи комерційною таємницею. Таке трапляється під час переписів населення, коли респонденти (частіше це одинаки, особи старшого віку чи такі, що мають великі статки) не погоджуються впускати обліковців з міркувань особистої безпеки (проникнення в помешкання підозрілих осіб під виглядом обліковців, повідомлення кримінальним елементам про майновий стан домогосподарства) або через занепокоєння можливістю передавання персональних даних про: дже-

рела засобів існування, місце роботи, вид зайнятості, місце перебування нелегальних мігрантів тощо).

На повноту охоплення також впливатиме невідповідність місця реєстрації та фактичного проживання членів домогосподарства. Особливої актуальності це набуває за сучасних умов високої мобільності населення (міграція, вимушене переселення, безпритульність або наявність кількох місць мешкання). Запобіганню таким ускладненням сприятиме широка й доступна інформаційно-роз'яснювальна компанія у засобах масової інформації, а також підтримка з боку громадськості і правоохоронних органів. Також на етапі підготовки перепису здійснюється попередній, так званий контрольний обхід обліковцями переписних ділянок для з'ясування відповідності наявних списків респондентів реальному стану, інструктаж населення про необхідність отримати відкріпний талон для осіб, які на момент перепису будуть відсутні за місцем реєстрації, тощо.

Друга особливість перепису — докладність обстеження. До програми переписів включають запитання, які дозволяють уточнити наявні дані (результати обліків, реєстрів, звітності), а також перелік запитань, що можна з'ясували лише безпосередньо спілкуючись з респондентом або під час огляду та підрахунку. Так, у процесі індивідуальної реєстрації кожної одиниці фіксуються всі її характеристики: формальні, що можна документально підтвердити, і неформальні, що відповідають реальному стану. Важливо, що під час перепису населення для встановлення фактів або їхнього підтвердження не вимагаються особові документи, а застосовується принцип самовизначення. Це сприяє довірливому спілкуванню і з'ясуванню особистісних питань на кшталт: фактичний шлюбний стан (зареєстрований/ незареєстрований шлюб або розлучення); самовизначення національності, рідної мови та володіння іншими мовами; робота в місцевості проживання/іншому населеному пункті або наміри і причини емігрувати з країни тощо). Інший приклад, коли переписом житлового фонду встановлюється не зареєстрований у технічному паспорті, а фактичний стан і тип окремо кожного житлового приміщення, його благоустрій та оснащення (водо-, газо-, електропостачання, наявність каналізації, системи видалення побутових відходів, доступ до мережі інтернет), розмір і реальна заселеність житлової площі, наявність земельної ділянки та сільськогосподарських тварин).

Третьою особливістю переписів є стислість термінів проведення. Це зумовлено не лише міркуваннями економії фінансових і людських витрат. Тривалість проведення перепису впливає на якість реєстрації. Дуже короткий період можливий у разі звуженої програми, короткого переліку простих запитань. Натомість, розтягнутий в часі перепис віддаляє реєстрацію подій та фактів від об'єктивного часу — критичного моменту. Наприклад, заповнення переписних листків під час перепису населення та житлового фонду зазвичай здійснюється впродовж 7–14 днів. Суб'єктивний час перепису може бути й довшим, якщо переписні дільниці розміщуються у важкодоступній місцевості (гірська, лісова тощо) або одиниці спостереження розпорочені на великій негусто заселеній території.

Із тривалістю переписів пов'язана ще одна їхня особливість — одномоментність. Як зазначалося у попередніх параграфах (1.3 та 2.1), реєстрація даних по всій території, об'єктах та одиницях спостереження має бути прив'язана до певного критичного моменту («дати перепису»). У переписах населення це не лише дата, а й година доби (наприклад, у ніч з 20 на 21 січня, тобто на 24–00), за станом на яку обліковуватимуться дані. Або, скажімо, перепис поголів'я сільськогосподарської худоби відбувається за станом на початок року (01.01). Це необхідно, щоб усі факти, події та одиниці спостереження були зафіксовані одночасно, оскільки впродовж суб'єктивного часу від початку до завершення перепису можуть відбуватися певні зміни із характеристиками одиниць спостереження.

Одночасність реєстрації важлива не лише для окремих об'єктів чи країн. Так, за сучасних умов небувалої активізації міжнародного міграційного руху населення (трудова, освітня, нелегальна міграція, неконтрольовані потоки біженців) дуже важливо скоротити розбіжності в часі проведення національних переписів. Зокрема, за рекомендаціями ООН щодо переписів населення та житлового фонду в раунді 2010 року, для 27 країн Євросоюзу був законодавчо встановлений 2011 рік. Раунд переписів 2020 року рекомендується ООН та Конференцією європейських статистиків провести в рамках Всесвітньої програми переписів 2015–2024 років.

Умова єдності та зіставності стосується не лише періоду і часу проведення перепису, а й застосування єдиної для всіх одиниць програми спостереження. Тобто облік одиниць здійснюється за однаковим переліком запитань. Якщо суцільний перепис здійс-

нюється у комбінації з вибірковою, то окремі питання мають заповнювати лише одиниці, які охоплюються певним процентом вибірки.

Забезпечення єдності та одномоментності було б неможливим, якби весь процес перепису не був централізованим. Усі види робіт, починаючи від підготовки до перепису (упорядкування переліку одиниць, їхнього фактичного розміщення, розроблення програми, інформаційна та роз'яснювальна кампанія, підготовка фахівців, залучення державних місцевих та відомчих адміністрацій) до власне проведення (реєстрація), контроль, нагромадження, опрацювання даних та поширення результатів — потребує централізованого керування разом з оперативним вирішенням питань на місцях.

Результати переписів мають багатоцільове використання впродовж тривалого періоду. Так, переписи населення — це єдиний надійний спосіб уточнення результатів поточного обліку населення; даних реєстрів населення та інших статистичних банків даних, пов'язаних із мешканцями території певної країни. Окрім того, переписи дають унікальну можливість створити якісну генеральну сукупність даних, з яких потім формуватимуться вибіркові сукупності для несучільних обстежень.

Якщо йдеться про перепис населення, то його результати — кількісні, якісні та структурні характеристики населення — використовують для розрахунку узагальнених показників соціально-економічного розвитку на макро-, мезо- і мікрорівнях країни. Дані перепису беруться як базові для здійснення перспективних розрахунків чисельності й складу населення загалом та в розрізі окремих територій і видів поселення, а також для проведення наукових досліджень. Матеріали переписів населення слугують інформаційним забезпеченням управління соціально-економічним розвитком країни та формування адекватної державної економічної й демографічної політики, використовуються для адміністративних цілей. Окрім того, результати національних переписів є невіддільними складовими всесвітнього банку статистичних даних з питань народонаселення, використовуються для міжнародних порівнянь та розроблення світових програм розвитку народонаселення.

Матеріали сільськогосподарського перепису слугують для визначення реального розміру господарств різної адміністративно-господарської та правової форми. Використовуються вони для уточнення даних Земельного кадастру, а також інформації про

форми володіння землею, землекористування, зрошення землі, поголів'я худоби за видами тварин. Окрема сфера застосування — визначення реальних обсягів трудових ресурсів сільського господарства та формування політики зайнятості і соціального захисту. На підставі даних сільськогосподарського перепису формується державна стратегія підтримки вітчизняного сільгоспвиробництва. Цей інформаційний ресурс використовується для оцінювання очікуваних обсягів сільськогосподарської продукції як з дослідницькою метою, так і для формування політики державної закупівлі сільськогосподарської продукції, оптимізації експортно-імпортних операцій тощо.

За результатами перепису об'єктів незавершеного будівництва визначається кількість та склад об'єктів незавершеного будівництва за характером і формою власності інвестора, адміністративно-правовим статусом забудовника й оцінюється ефективність інвестиційних проектів, ефективність роботи компаній-збудовників. Окрім того, визначається стратегія житлового кредитування та державних субсидій; оцінюються перспективи покращення житлових умов, розвитку інфраструктури. Вирішуються питання про введення в дію об'єктів господарського призначення; оцінювання обсягу необхідних робіт та окупності витрат для завершення будівництва.

Утім, попри всю унікальність перепису як суцільного виду спостереження і незамінного джерела докладної, всеосяжної і надійної інформації вони мають певні обмеження порівняно з іншими формами та видами спостереження.

По-перше, це складність організації і проведення широкомасштабного спостереження у стислі терміни.

По-друге, висока фінансова витратність. Близько 50–60 % витрат іде на підрахунок та облік, який потребує великої «армії» тимчасово залучених та підготовлених обліковців (витрати на навчання, оплату праці та на проведення пробного перепису). Чималих витрат потребують такі підготовчі роботи, як розроблення програми, складання плану, схеми та карти розміщення одиниць спостереження, розроблення статистичного інструментарію та макетів вихідних таблиць, розроблення програмного забезпечення: кодування і перенесення даних на електронні носії, контролю, статистичного опрацювання. Окремих витрат потребують логістика, друк бланків, контроль і опрацювання даних, систематизація та гармонізація з іншими формами спостереження, публікація і розміщення результатів.

По-третє, це нерівномірність розподілу в часі матеріальних, фінансових і трудових витрат. На короткий період реєстрації даних слід мобілізувати чималі інформаційні, матеріальні та людські ресурси. Очевидно, що максимальні витрати всіх ресурсів припадають на рік проведення перепису. Оскільки переписи проводяться періодично зі значним часовим інтервалом, додатковою проблемою стає забезпечення постійних якісно підготовлених кадрів, які б могли зорганізувати і проводити перепис, застосовуючи та передаючи свої знання та досвід, набутий у попередньому переписі.

По-четверте, повнота й об'єктивність реєстрації під час перепису значною мірою залежить від згоди і бажання респондентів.

Наостанок слід взяти до уваги значний розрив між часом проведення переписів і оприлюдненням їхніх результатів. Уже зрозуміло, що перепис не закінчується у день, коли збігає суб'єктивний час. Залежно від рівня технічного оснащення процедури обліку (паперові чи електронні носії), значно варіює тривалість процесу передавання даних від обліковців до інструкторів, а від них — керівникам територіальних переписних відділів. Після цього розпочинається величезна робота з перевірки даних реєстрації, кодування, нагромадження. Далі розробляються вихідні таблиці, здійснюється зведення і групування даних, розрахунок узагальнених показників, опис та аналіз результатів опрацювання, розміщення даних на паперових та електронних ресурсах, оприлюднення результатів. Окрема величезна робота — це актуалізація результатів перепису станом на початок року, перерахунок даних поточного обліку за міжпереписний період тощо. З огляду на те, що результати переписів проходять тривалий шлях від моменту збирання статистичних даних до моменту оприлюднення узагальнених матеріалів, інформація може частково втрачати свою актуальність.

Особливості сучасних переписів: міжнародний і вітчизняний досвід

Історія переписів бере початок ще до нашої ери. Держави давньої цивілізації збирали дані про чисельність, статево-віковий та майновий склад населення, розмір землі та майна. Існують відомості про переписи населення Азії, та Африки від 3 тис. до н. е.:

Єгипет (2800–2250 роки до н. е.); Китай (2238 р. до н. е.); Давній Рим (із 435 р. до н. е.) — регулярні цензи римських громадян за місцем їхнього проживання (ім'я, вік, майновий стан).

За часів Середньовіччя та панування феодального ладу (від V до середини XVII століття) роздрібнені держави систематичних обліків населення на загальнодержавному рівні не проводили. Здійснювалися лише разові реєстрації населення, яке залишилося в живих по закінченні міжусобних війн та епідемій. На початку XVIII століття почали відновлюватися загальнодержавні обліки населення, які охоплювали лише платників податків і тих, хто підлягав рекрутським наборам. Приміром, за царювання Івана Грозного у Московії для податного обкладання проводилися *поземельні обліки*, а у XVII столітті одиницею податного обкладання стає «двір» (господарство).

Перший перепис населення в сучасному розумінні був проведений у США 1790 року. Згодом відбуваються переписи в європейських країнах (Фінляндія, Швеція — 1800 рік), (Англія, Франція, Данія — 1801 рік).

На території сучасної України теж збиралися дані про населення, зокрема *Скіфського Царства* — облік чоловічого населення (за кількістю мідних стріл); *Київської Русі* (IX ст.) — облік податі. За часів татарської навали на території Русі обліки населення проводилися з метою обкладання дворів податками. А за період 1255–1259 років проводився подушний облік окремих верств населення.

У *Російській імперії* за наказом Петра I (26.11.1718 року) запроваджено «подушний облік» осіб чоловічої статі, які обкладалися податками і підлягали рекрутським наборам. Дворянство, державні чиновники не обліковувалися, а для решти категорій записи особових даних заносилися до «подушних списків». Перші подушні списки були складені у 1720–1721 роках і називалися «ревизькими сказками». Такі «ревізії» тривали майже півтора століття і востаннє проводилися у 1860 року.

Повноцінним переписом, за великою науково обґрунтованою програмою, був *Перший Всеросійський перепис населення* 1897 року. Він включав запитання, що з'ясовували стать, повний вік, відношення до глави сім'ї, сімейний стан, соціальний статус, місце народження і приписки, а також постійного мешкання, віросповідання, рідну мову, грамотність, вид занять. Ця програма згодом використовувалася для підготовки подальших переписів.

За радянських часів найбільш вдалим, повним і добре організованим був суцільний перепис наявного населення і житлових умов 1926 року (без урахування сучасної території Західної України). Населення, яке мешкало на території теперішніх Волинської, Івано-Франківської, Львівської, Рівненської, Тернопільської, Закарпатської і Чернівецької областей, було зареєстровано польськими, румунськими і чеськими переписами 1921, 1930 і 1931 років. Наступні два Всесоюзні переписи 1937 і 1939 років стосувалися лише населення (без обстеження житлового фонду) і проводилися за дуже вузькою програмою. Результати обох переписів не публікувалися, в тому числі перепису 1937 року з так званими «дефектними» матеріалами, що не підтверджували задеклароване зростання чисельності радянського населення. За повоєнної доби Всесоюзні переписи населення проводилися тричі — у 1959, 1970, 1979 роках, і вони теж не враховували житлових умов, але стосувалися не лише наявного, а й постійного населення, і проводилися у комбінації з вибірковою обстеженням (25 % вибіркового перепису). Програма останнього Всесоюзного перепису населення 1989 року вперше після 1926 року включала обстеження житлового фонду і проводилася за схемою суцільного і вибіркового обстеження.

Перший Всеукраїнський перепис населення 2001 року організувався лише на суцільній основі. Крім соціально-економічних і демографічних характеристик наявного і постійного населення, в переписному листку (Додаток Д) фіксувалися: тип і розмір житлового приміщення, кількість зайнятих кімнат, а також наявність і розмір земельної ділянки сільськогосподарського призначення. Питання комфортності житла не з'ясувалися.

Упродовж наступного десятиліття світовий інформаційний простір та його технічне забезпечення зазнали кардинальних змін, які висунули нові вимоги до сучасних підходів до організації переписів. Досвід проведення Всесвітнього перепису населення і житлового фонду раунду 2010 року показав не лише можливість, а й доцільність пошуку альтернативних традиційному методу проведення сучасних переписів. Так, у раунді переписів 2010 року порівняно із попереднім 2000 року частка країн, в яких населення переписувалося традиційним методом, зменшилася з 80 % (40 із 50 країн) до 56,4 %. Дедалі поширенішими стають комбіновані методи та переписи на базі реєстрів. У раунді 2010 року чверть (25,4 %) європейських країн провели перепис за

змішаною схемою (реєстри та вибіркові обстеження домогосподарств), а вісім країн (14,5 %) користувалися виключно даними адміністративних реєстрів.

Раунд переписів 2020 року у світі проводитиметься за принципово новими методологічними підходами з огляду на можливість комплексного й оперативного використання наявних різноманітних електронних банків даних про населення. Так, персональні дані, що містяться і повторюються в кількох адміністративних реєстрах, можуть комбінуватися з додатковою інформацією в результаті спеціальних вибірових обстежень.

Запровадження альтернативних форм перепису населення зумовлене кількома обставинами:

- економією фінансових і людських витрат;
- зміною запитів з боку користувачів інформації щодо її своєчасності, актуальності й різноплановості;
- небажанням населення співпрацювати, з одного боку, та зниженням навантаження на респондентів — з іншого.

Зазначимо, що саме величезні бюджетні витрати впродовж короткого терміну (7–14 днів), необхідні для залучення, підготовки й організації роботи інтерв'юєрів, а також для реєстрації і контролю даних, опрацювання та публікації матеріалів перепису спонукають до пошуку менш витратних методів і прийомів проведення переписів. Останнім часом ширшого застосування набуває заповнення респондентами переписного листка через мережу інтернет. Зокрема, у 2011 році таким способом проводився перепис населення у 13-ти країнах світу, і таким чином було суттєво заощаджено кошти.

Не слід забувати також, що результати переписів проходять тривалий шлях від моменту збирання статистичних даних до моменту оприлюднення узагальнених матеріалів. За цей час інформація частково втрачає свою актуальність, особливо стосовно динамічних явищ і процесів (зміна форми власності житлових об'єктів або місця проживання чи виду діяльності людей).

Чималу роль у виборі інших форм відіграє небажання респондентів брати участь у переписах. Дедалі частіше трапляються випадки, коли населення відмовляється відчиняти двері обліковцям з міркувань безпеки або протестує проти багаторазового запиту їхніх персональних даних.

З огляду на всі зазначені обставини в сучасній міжнародній практиці поширюються модифікації традиційного перепису

(рис. 3.2). Зокрема, в Канаді та США (до 2010 року) застосовують комбінацію суцільного обстеження за коротким переписним листком (основні демографічні характеристики) і вибіркового за деталізованою програмою (довгий переписний листок). Перевагою цього методу є розширення переліку різноманітних запитань при одночасному зниженні навантаження на респондентів і скороченні витрат на збирання й опрацювання даних. Під час перепису населення Сполучених Штатів Америки 2010 року запроваджено видозмінену методику: комбінацію суцільного перепису за коротким переписним листком (раз на 10 років) із докладним щорічним вибіркоvim обстеженням за довгим листком.

Інший підхід змішаного перепису — комбінація даних адміністративних реєстрів із суцільним обстеженням шляхом прямого опитування. Порівняно із традиційним переписом такий метод суттєво зменшує навантаження на респондентів, оскільки частина особових даних береться з реєстрів і уточнюється з огляду на повноту обліку та актуальність інформації. Під час суцільного обстеження з'ясовуються індивідуальні дані, що не обліковуються реєстром. Його застосовували у 2011 році колишні радянські країни Балтії, Чеська Республіка, Італія та Іспанія.

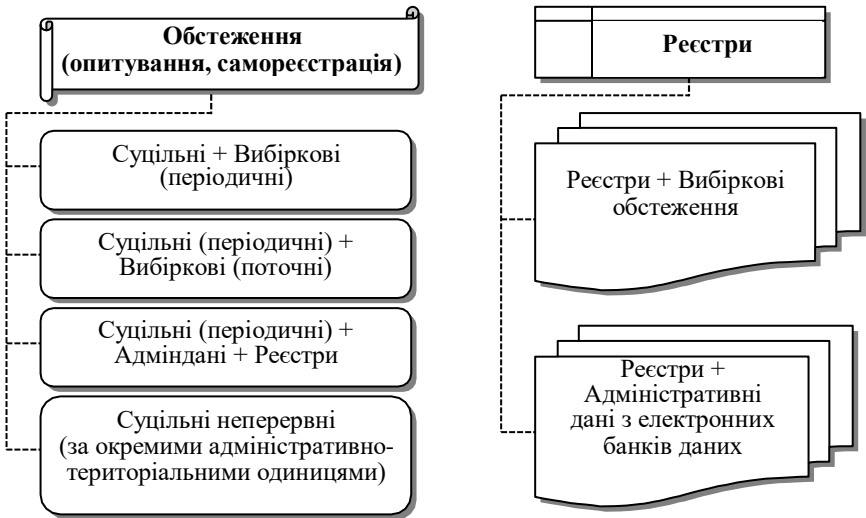


Рис. 3.2. Комбіновані моделі сучасних переписів населення

Ще один варіант комбінації реєстрів із різноманітними вибірковими обстеженнями, що проводяться на регулярній основі, застосовують Нідерланди і Словенія. Але такий прийом потребує жорсткої ув'язки записів персональних даних по кожному респонденту, що беруться з різних інформаційних ресурсів.

Цікавий досвід проведення неперервного перепису існує у Франції (від 2004 року). Неперервність забезпечується методом ковзної середньої із п'ятирічним інтервалом. Тобто серед нечисленних адміністративно-територіальних одиниць, що налічують до 10 тис. осіб населення, щороку організується суцільне обстеження однієї із п'яти груп населених пунктів. У великих населених пунктах щорічно проводиться восьмивідсоткове вибіркове обстеження домогосподарств. Таким чином, упродовж п'яти років охоплюється переписом 70 % усього населення Франції, а його результати не втрачають актуальності та щорічно публікуються.

Принципово інший підхід — перепис населення винятково на підставі реєстрів. Уперше в світі таку методiku застосувала Данія (1981 р.), а згодом до неї приєдналися скандинавські країни: Фінляндія (1990 р.), Швеція (2005 р.), Норвегія (2011 р.), а також Австрія, Бельгія і Словенія. За цією методикою у першу чергу використовують дані реєстру населення, де по кожному мешканцю території окремого муніципалітету міститься особова інформація (прізвище, ім'я, стать, дата і місце народження, місце мешкання, громадянство, сімейний стан), а також роль в сім'ї. Можуть додаватися відомості про вид діяльності, рівень освіти тощо. Такі реєстри складають за даними попередніх переписів населення, а в подальшому постійно оновлюють. Актуалізація здійснюється шляхом обов'язкового особистого повідомлення (зміна місця мешкання, виду діяльності). Персональні дані цього реєстру автоматично з'єднуються з даними інших загальнодержавних реєстрів (сімей, освіти, зайнятості, фізичних осіб-суб'єктів господарювання) або з картотеками (соціального та медичного страхування, нерухомості тощо) та даними поточного обліку (народження, смерті, шлюби, розлучення). Наявність взаємозв'язаних реєстрів та електронних банків даних різних відомств дозволяє безперешкодно, навіть і щорічно, здійснювати облік населення без залучення додаткових фінансових і людських ресурсів і зайвого навантаження на респондентів.

Утім, реєстри та різні банки даних не можуть гарантувати повноту охоплення всього, навіть законослухняного населення, яке мешкає на території держави. Реєстри також не містять необхідної інформації мотиваційного характеру (причини незайнятості, мотиви зміни місця проживання, еміграції тощо) або етнолінгвістичних та майнових даних (джерела засобів існування). На підставі даних реєстрів не можна оцінити реальні масштаби міграції, статево-віковий склад мігрантів, а без цього неможливо визначити чисельність наявного і постійного населення за статтю, віком і видом поселення. Таку інформацію одержують лише в процесі опитування респондентів. Також важливою проблемою є захист персональних даних від несанкціонованого доступу під час агрегування або використання особистих відомостей з різних інформаційних ресурсів.

Узагальнюючи міжнародний досвід проведення переписів, слід наголосити на положеннях нової концепції Всесвітнього перепису населення і житлового фонду раунду 2020 року:

- перехід від традиційної моделі перепису до комбінованої — з одночасним залученням різних інформаційних ресурсів (адміністративних даних, суцільних та вибіркового спеціально організованих обстежень);

- перехід від схеми періодичних переписів (раз на 10 років) до неперервних (ковзних) переписів;

- розширення переліку тематичних питань програми перепису завдяки приєднанню даних різноманітних спеціально організованих обстежень;

- комплексне використання результатів різних переписів;

- широке запровадження сучасних способів реєстрації первинних даних (електронна реєстрація в мережі інтернет);

- синхронізація проведення переписів населення в різних країнах та міжнародна зіставність їхніх результатів.

Реалізація цієї концепції дає змогу суттєво скоротити бюджетні витрати, знизити навантаження на респондентів та персонал перепису, підвищити актуальність результатів, удосконалити порядок ведення реєстрів (оновлення і контроль даних). Водночас новий підхід висуває нові виклики щодо технічного і програмного забезпечення узгодженого збирання, нагромадження, захисту і використання персональних даних з різних інформаційних ресурсів.

Перепис населення: етапи проведення, програмно-методологічні та організаційні питання

Найточніші дані про природний та механічний рух населення, його чисельність і склад одержують під час переписів населення, які дозволяють скоригувати, уточнити дані поточного обліку.

Перепис населення — суцільне спостереження всіх без винятку мешканців регіону за для визначення їхньої чисельності та складу станом на певну дату. *Метою* перепису населення є одержання об'єктивної інформації про зміни, що відбулися в соціально-економічному житті та устрої держави від часу попереднього перепису; створення інформаційної бази демографічних та соціально-економічних даних щодо чисельності і складу населення в країні загалом та за адміністративно-територіальними одиницями.

Сучасний перепис здійснюється за єдиним затвердженим планом, що включає: організацію і проведення підготовчих робіт, методологічне розроблення перепису, безпосередньо реєстрацію даних, їхній контроль і оброблення, аналіз та публікацію і поширення результатів.

Перепис за традиційною методологією включає такі підготовчі роботи:

- уточнення меж міських поселень та великих сільських населених пунктів і складання актуального картографічного матеріалу;
- упорядкування адресних списків домогосподарств міських поселень та великих сільських населених пунктів;
- забезпечення повноти охоплення і коректності поточного обліку одиниць спостереження за місцем перебування;
- переписне районування — поділ території міських населених пунктів, районів, сільських населених пунктів на переписні дільниці із закріпленою за ними кількістю переписного персоналу, засобів матеріально-технічного забезпечення, переписного інструментарію тощо;
- добір і навчання переписного персоналу (обліковців, інструкторів-контролерів, начальників переписних відділів, операторів);
- проведення пробного перепису;

- підготовка технічного і програмного забезпечення для введення, контролю, узагальнення, опрацювання, нагромадження та зберігання результатів перепису.

Перед проведенням перепису населення розробляють програмно-методологічні питання, які передбачають визначення об'єкта спостереження, одиниці сукупності, одиниці спостереження, розроблення програми перепису, статистичного інструментарію та процедури контролю.

У більшості переписів населення *об'єктом спостереження* є постійне та наявне населення з урахуванням тимчасово присутніх і тимчасово відсутніх осіб упродовж 12 місяців, що передували перепису. *Одиницею сукупності* перепису є фізична особа, яка мешкає на території, де здійснюється перепис. Згідно з Програмою Всеукраїнського перепису населення 2001 та 2020 років одиницею сукупності є кожен мешканець України (громадянин України або іншої держави та особа без громадянства), незалежно від наявності місця реєстрації респондента, його майнових прав на житло, в якому він фактично мешкає.

Обома переписами охоплюються одночасно три *одиниці спостереження*: окрема людина, сім'я і домогосподарство.

Сім'я — сукупність людей, які проживають спільно й пов'язані шлюбом, кровною спорідненістю чи свояцтвом, а також спільним побутом і взаємною відповідальністю за дітонародження і соціалізацію дітей та підтримку членів сім'ї.

Під час перепису з'ясовується модель сім'ї (повна, неповна, проста) за допомогою запитання про родинні стосунки із особою, записаною першою. Сімейний склад населення обстежується з метою визначення умов для відтворення населення, шлюбної ситуації, соціальної політики захисту багатодітних сімей, однаків, а також як інформаційна база для бюджетних обстежень з урахуванням розміру і складу сімей.

Домогосподарство — сукупність людей, які проживають спільно в одному житловому приміщенні або його частині, ведуть спільне господарство, повністю або частково об'єднують та витрачають кошти.

Члени домогосподарства можуть перебувати в родинних взаєминах або стосунках свояцтва, не перебувати у будь-яких з цих стосунків або бути і в тих, і в інших стосунках. Домогосподарство на відміну від сім'ї може складатися з однієї особи. Дані про структуру домогосподарств використовуються не лише для роз-

роблення програми соціальної підтримки та адресної допомоги уразливим групам населення. Їх беруть до уваги при аналізі бюджетів домогосподарств (доходів і витрат), структури і характеру споживання та заощаджень, економічної активності населення, оцінки житлових умов тощо.

Важливими складовими програмно-методологічних питань плану перепису є програма спостереження і статистичний інструментарій. Програму Всеукраїнського перепису населення затверджено наказом Держстату України у відповідності до ст. 5 Закону України «Про проведення Всеукраїнського перепису населення» та Рекомендації Конференції європейських статистиків щодо проведення перепису населення і житлового фонду 2010 і 2020 років.

Програма перепису — перелік запитань, за якими обстежується населення. Програма Всеукраїнського перепису населення України 2001 року включала адресні відомості (населений пункт, адресу, прізвище, ім'я та по батькові, відношення до члена домогосподарства, записаного першим), а також особові запитання (стать, вік, дату і місце народження, сімейний стан, національність, громадянство, рідну мову та мову, якою вільно володіють). Також містилися запитання про рівень освіти, тип навчального закладу, джерела засобів існування, вид зайнятості та діяльності, професію та посаду. Жінок віком 15 років і старше запитували про кількість народжених дітей, у тому числі живих.

Програму перепису 2020 року доповнено запитаннями щодо відвідування дітьми дошкільних закладів, міграційної й економічної активності населення (зайнятості і причини безробіття та вимушеної міграції, мотиви еміграції). Включено запитання про стан житлових умов населення, характер благоустрою житлового приміщення із докладним переліком наявних комунальних послуг включно з інтернетом, час побудови і форму власності житлового приміщення; наявністю сільськогосподарських тварин.

За програмою Всеукраїнського перепису населення України 2001 року застосовувався такий інструментарій: форма 1 «Список проживаючих у приміщенні членів домогосподарства № _ та їхні житлові умови», форма 2с «Переписний лист» (Додаток 1, 2), форма 3 «Переписні документи на приміщення», а також набір інструктивних матеріалів щодо організації та проведення перепису.

Форма 1 заповнювалася для домогосподарства в цілому. В ній окрім ідентифікаційної частини (номер переписного відділу, інструкторської та лічильної дільниці, номер приміщення) та адресної частини (область, населений пункт, адреса, прізвище домовласника), містилося два розділи. Перший розділ — поіменний список мешканців — членів домогосподарства (для тимчасово відсутніх: причина і тривалість відсутності). Другий розділ стосувався житлових умов домогосподарства (тип житла; кількість зайнятих житлових кімнат; розмір житлової та загальної площі; наявність земельної ділянки сільгосппризначення, її розмір, право власності та перебування в оренді).

Форма 2с заповнювалася персонально на кожного мешканця із зазначенням прізвища, ім'я та по-батькові й 19 запитань, що включені до Програми перепису (див. вище). У формі зазначається, що записи стосовно респондента не підлягають поширенню без його згоди і використовуватимуться лише для статистичних цілей у зведеному, деперсоніфікованому вигляді.

Форма 3 «Переписні документи на приміщення» заповнювалася для кожної окремої квартири, будинку або його частини, кімнати у гуртожитку. Для населення, яке перебувало в інституціональних установах (дитячих будинках, школах-інтернатах, інтернатах, монастирях) та в лікарнях, санаторіях, готелях тощо, форма заповнювалася на окремий корпус, клас, палату, номер. У ній зазначалася ідентифікаційна та адресна частини приміщення. Також містилось три допоміжні таблиці із зазначенням кількості зареєстрованих респондентів та переліку заповнених на них документів.

Традиційний перепис населення здійснюється на суцільній основі *експедиційним способом* або *самореєстрацією*. Всеукраїнський перепис населення 2001 року проводився експедиційним способом, тобто переписний бланк заповнювався підготовленим інтерв'юером (обліковцем) зі слів респондентів. У деяких країнах населення переписується шляхом самореєстрації за бланками, надісланими поштою (США), а решта неохоплених одиниць реєструється обліковцями.

Стрімкого поширення набуває спосіб заповнення шляхом самореєстрації електронних переписних листків через мережу інтернет. Уперше такий експеримент провела Іспанія під час перепису 2001 року. Згодом у раунді переписів 2010 року в такий спосіб

переписувалося населення США, Канади, Бразилії, Австралії, Японії, Швейцарії та Естонії.

Безперечними перевагами цього способу є: оперативність збирання даних; полегшений доступ до мешканців мегаполісів, а також віддалених та малих територій; економія фінансових і людських ресурсів; можливість здійснювати моніторинг кількості та якості заповнених переписних бланків в онлайн режимі.

Міжнародний досвід Інтернет-перепису раунду 2010 року показав, що таким чином забезпечується 25–60% охоплення респондентів. Ефективність запровадження такого способу реєстрації вочевидь залежить від розгалуженості мережі інтернет, якості й безперервності зв'язку, а також від готовності респондентів до відповідального заповнення електронних документів.

Інша новація у способі реєстрації — заповнення обліковцем електронної форми переписного листка зі слів респондента на планшетній комп'ютер. Така процедура у раунді 2010 року застосовувалася в Польщі та Ізраїлі. Завдяки цьому вдалося виключити паперові носії переписних листків і пришвидшити процес зведення і опрацювання даних перепису. Це також дозволило заощадити не лише матеріальні, а й трудові витрати, пов'язані з перенесенням даних з одних носіїв на інші, кодуванням даних, транспортуванням і збереженням бланків. Водночас істотно підвищилася якість реєстрації завдяки уникненню помилок при перенесенні даних з паперових на електронні носії.

Під час Всеукраїнського перепису населення України 2001 року застосовувалися паперові бланки, а перенесення даних на електронні носії після контролю і кодування здійснювалося шляхом сканування. За допомогою пристроїв оптичного зчитування та спеціального програмного забезпечення виконувалось: автоматичне розпізнавання переписних документів, первинний контроль даних, формування текстових файлів, запис відсканованих графічних образів переписних документів.

Але запровадження електронної реєстрації потребує відповідного технічного оснащення, програмного забезпечення та підготовлених обліковців.

Традиційна модель перепису населення передбачає періодичні обстеження, як правило, через однакові проміжки часу (за рекомендаціями ООН не частіше ніж раз на 10 років, у роки, що закінчуються на «9» або «0»). Перепис виконується за місцем

постійного проживання населення. Якщо особа на момент перепису відсутня, то її реєструють як тимчасово відсутню.

Для проведення перепису населення встановлюється *критичний момент*, станом на який реєструється населення. Так, за програмою Всеукраїнського перепису населення України 2001 року, критичним моментом був час 24.00 із 4-го на 5-е грудня. Оскільки переписати все населення неможливо в одну мить, то встановлюється період спостереження, який у 2001 році тривав 10 днів (5.12–14.12).

Проведенню перепису населення передують пробний перепис, під час якого в реальних умовах тестуються Програма та статистичний інструментарій. Також відпрацьовуються організаційні питання, зокрема перевіряється використання під час перепису планшетних комп'ютерів і ноутбуків, а також програмне забезпечення внесення статистичної інформації та її опрацювання. Розпорядженням Кабінету міністрів України Держстат має провести пробний перепис населення Оболонського району м. Києва та Бородянського району Київської області у грудні 2019 року.

Заповнені обліковцями переписні документи проходять процедуру підготовки для подальшої статистичної обробки даних. По-перше, здійснюється попередній контроль записів на повноту і коректність шляхом перевірки за допомогою відповідних словників-класифікаторів (національностей і мов, видів діяльності, освітніх рівнів). Далі звіряється кількість заповнених форм 1 з числом домогосподарств, зазначених у формі 3, а також кількість заповнених форм 2с з числом членів домогосподарства. Уточнюються і перевіряються адресні дані переписних листків на їхню відповідність даним форм 1 і 3. Перевірені дані кодуються і переносяться на електронні носії, якщо переписні документи були в паперовому вигляді.

Після контролю перенесених даних формуються електронні банки даних. Для подальшого опрацювання і систематизації даних розробляють макети вихідних таблиць: простих, групових і комбінаційних. Зведені і згруповані дані аналізуються, порівнюються в часі та за окремими територіями чи видами поселення. Попередні результати в узагальненому вигляді оприлюднюються.

Окремий етап роботи — це перерахунок даних поточного обліку за матеріалами періодичних переписів та побудова неперервних динамічних рядів за міжпереписний період.

Результати перепису населення використовуються для актуалізації, доповнення й уточнення інших форм статистичного спостереження:

- узгоджуються з даними поточного обліку (документальний облік або статистична звітність). Оцінюються та аналізуються розбіжності, з'ясовуються причини розбіжностей;
- беруться за основу для формування або уточнення статистичних реєстрів;
- використовуються як база для проведення моніторингів та спеціально організованих вибіркового обстежень;
- застосовуються для складання картотек, кадастрів;
- використовуються як основа для проведення референдумів.

Статистичні реєстри та формування статистичних банків даних

Статистичний реєстр — список або перелік одиниць певного об'єкта спостереження із зазначенням необхідних ознак, який складається та оновлюється під час постійного відстежування.

Статистичні реєстри пройшли тривалий шлях еволюції — від первинних форм (поіменних записів, списків та картотек) до сучасних складних електронних інформаційних систем, інтегрованих з іншими інформаційними ресурсами різного рівня централізації. За доби стрімкого розвитку цифрової (електронної) економіки та глобалізації суспільних процесів статистичні реєстри виконують роль не лише нагромаджувача інформації, а й своєрідного комунікатора між різними інформаційними ресурсами (див. рис. 3.3).

У сучасному тлумаченні *статистичний реєстр* — це електронна інформаційно-телекомунікаційна система, призначена для збирання, накопичення, захисту, опрацювання, поширення, використання та зберігання даних щодо респондентів статистичних спостережень.

До ідентифікаційних ознак реєстрів належать ідентифікаційні коди та реєстраційні номери одиниць спостереження. Наприклад, у *Реєстрі статистичних одиниць*, що використовується в Державному статистичному структурному спостереженні, такими

ознаками є ідентифікаційний код суб'єкта господарювання відповідно до реєстрації юридичних осіб в ЄДРПОУ; індивідуальний код місцевої одиниці, найменування і юридичний статус суб'єкта.

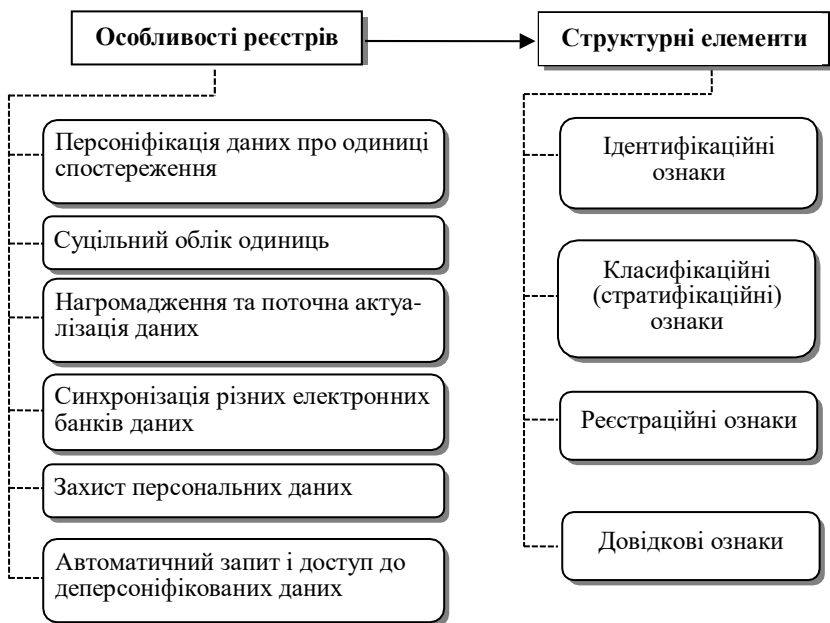


Рис. 3.3. Особливості статистичних реєстрів та їхні структурні елементи

У Державному реєстрі речових прав на нерухоме майно та їх обтяжень автоматично формується та надається реєстраційний номер об'єкта нерухомого майна, реєстраційний номер заяви, запису, індексний номер рішення про реєстрацію прав та їх обтяжень. Ідентифікатором об'єкта нерухомого майна є його адреса. Він складається з назви адміністративно-територіальної одиниці, району, населеного пункту або селищної чи сільської ради, назви вулиці (проспекту, бульвару, площі, провулку тощо), номера будинку, будівлі, споруди, номера окремої частини об'єкта нерухомого майна (квартири).

Стратифікаційні ознаки дають сутнісну характеристику одиниці спостереження. Так, для суб'єктів господарювання зазначається вид економічної діяльності за КВЕД; організаційно-правова форма господарювання за КОПФГ; середня кількість зайнятих, у тому числі найманих працівників тощо. Наприклад, у *Реєстрі статистичних одиниць АГРО* (у сільському господарстві, мисливстві, лісовому і рибному господарстві) поряд із вище згаданими ознаками вказується земельна площа, в тому числі сільськогосподарських угідь, з них узята в оренду площа ріллі, чисельність сільськогосподарських тварин за основними видами. Або, повертаючись до прикладу *Державного реєстру речових прав на нерухоме майно*, стратифікаційними ознаками є: тип об'єкта нерухомого майна (підприємство, житловий будинок, квартира, земельна ділянка тощо); тип суб'єкта речового права (юридична чи фізична особа) із реквізитами документа, що посвідчує особу (реєстраційний номер облікової картки платника податків або унікальний номер запису в Єдиному державному демографічному реєстрі).

Реєстраційні ознаки вказують на дату первинної реєстрації та дату останніх реєстраційних змін. Наприклад, це може бути дата реєстрації суб'єкта господарювання або дата і час державної реєстрації набуття, зміни чи припинення права власності та інших речових прав на нерухоме майно, взяття на облік безхазяйного майна.

У реєстрах вказуються також довідкові і технологічні ознаки. Наприклад, для суб'єктів господарювання можуть вноситися дані про: кількість відокремлених підрозділів та їхній правовий статус; кількість місцевих одиниць; ідентифікаційний код ЄДРПОУ і найменування головного підприємства; телефон, прізвище, ім'я, по батькові керівника підприємства; дані щодо економічної та профільної активності суб'єкта тощо.

Утім, статистичні реєстри не слід розглядати лише як накопичувачі даних, якими користуються відповідні державні установи. Респонденти статистичних реєстрів також мають право отримати необхідну інформацію в межах, передбачених регулювальними законодавчими актами. Наприклад, з архіву Державного реєстру речових прав на нерухоме майно можна отримати дані про суб'єктів прав власності на об'єкти нерухомого майна із зазначенням їхніх реквізитів. Також є можливість ознайомитися зі станом заяв щодо реєстрації прав або судових

рішень стосовно заборони реєстрації чи її зняття. Окрім того, на веб-порталі Міністерства юстиції України періодично оприлюднюється статистична інформація про кількість і склад за видами проведених реєстраційних дій упродовж відповідного періоду.

Сьогодні в Україні функціонує розгалужена мережа державних реєстрів, які дозволяють вести постійний облік одиниць спостереження в різноманітних сферах суспільного життя. Наприклад, *Державний реєстр фізичних осіб — платників податків* створений для єдиного державного обліку фізичних осіб, які згідно із законом України (Кодексом та іншими нормативно-правовими актами) мають сплачувати податки та збори. Цей облік здійснюється з метою контролю правильності податкового нарахування, своєчасності і повноти сплати податків, нарахованих фінансових санкцій, дотриманням податкового законодавства.

У даному реєстрі одиницями спостереження є фізичні особи — платники податків: громадяни України; іноземці та особи без громадянства, які постійно мешкають в Україні; іноземці та особи без громадянства, які не мають постійного місця проживання в Україні, але відповідно до законодавства зобов'язані сплачувати податки в Україні або є засновниками юридичних осіб, створених на території України.

Облік ведеться за реєстраційними номерами облікової картки платника податків. Дані стосовно осіб, які через релігійні переконання відмовилися отримати реєстраційний номер облікової картки, але офіційно про це повідомили, заносяться в окремий реєстр за прізвищем, ім'ям, по батькові (за наявності), серією та номером паспорта. Поточна реєстрація даних здійснюється на підставі податкової декларації про майновий стан і доходи, яка може подаватися в один з трьох способів. Наприклад, особисто платником у відділенні податкової інспекції або надсилатися поштою у паперовому вигляді чи електронною поштою із зазначенням електронного цифрового підпису.

Для реєстрації даних щодо податків та зборів різних фізичних осіб встановлюється диференційований календар спостережень. Так, для платників податку на доходи фізичних осіб — критичним моментом є 1 травня у наступному за звітним роком. Стосовно платників податку на доходи фізичних осіб — підприємців

суб'єктивний час спостереження — упродовж 40 календарних днів після закінчення звітного (податкового) року.

Важливо, що респонденти Державного реєстру фізичних осіб — платників податків можуть зробити запит і отримати в електронному вигляді особисті відомості про суми виплачених доходів та утриманих податків. Вільний доступ до власних даних забезпечується за допомогою клієнтського програмного забезпечення «Єдине вікно подання електронної звітності», яке розміщене на офіційному веб-сайті Державної податкової служби України.

Разом з Державним реєстром платників податків одним із перших за часів незалежності України був створений реєстр виборців спочатку у формі інформаційно-аналітичної підсистеми «Списки виборців», а вже від 2009 року повноцінно запрацював Державний реєстр виборців.

Державний реєстр виборців (ДРВ) — автоматизований інформаційно-телекомунікаційний банк даних призначений для обліку, зберігання, опрацювання, оновлення і нагромадження відомостей стосовно громадян України, які мають право голосу. Відповідно до статті 70 Конституції України таке право надається особам правоздатного віку (18 років і старші). Реєстр дозволяє постійно уточнювати чисельність та віковий склад виборців, адресу проживання, зміну імені та прізвища виборців і тим самим уникати помилок при написанні їхніх особових реквізитів. Внесення даних до реєстру здійснюється районним відділом ведення Державного реєстру виборців, який отримує відомості про громадян правоздатного віку від районних, міських та селищних органів місцевого самоврядування, органів Міністерства внутрішніх справ, Міністерства юстиції (РАЦСи), від військових частин (про військовослужбовців строкової служби), виправних колоній, органів системи охорони здоров'я (стосовно осіб, нездатних самостійно пересуватися), органів опіки та піклування (стосовно недієздатних осіб).

Зі згаданими вище реєстрами безпосередньо пов'язаний реєстр населення. *Реєстр населення* — поіменний список мешканців країни, який регулярно переглядається, поповнюється і містить відомості стосовно кожної особи, що постійно мешкає на території держави. У першу чергу це демографічні характеристики (паспортні дані та події, що відбуваються впродовж життя). У реєстрі можуть фіксуватися біометричні дані (завви-

чай відбитки пальців), а також деякі економічні параметри (доходи та сплата податків). Реєстри населення, як правило, складаються за матеріалами перепису населення і дедалі частіше використовуються замість поточного обліку. Ними користуються при складанні списків військовозобов'язаних, виборців, платників податків, а також у разі запровадження безпаспортного режиму.

Перші реєстри населення з'явилися у Швеції (1847), згодом у Бельгії та Нідерландах. На сьогодні реєстри населення створено практично в усіх країнах ЄС. Вони містять персональні дані всіх мешканців (прізвище, ім'я, стать, місце і дата народження, місце проживання, сімейний стан, роль щодо глави сім'ї, громадянство, вид діяльності тощо). Ці дані поповнюються новими демографічними подіями (народження, смерть, одруження, розлучення, зміна виду діяльності та місця мешкання) завдяки взаємопов'язаним електронним банкам даних різних інституцій.

В Україні від 2012 року чиним є *Єдиний державний демографічний реєстр* — це електронна інформаційно-телекомунікаційна система, призначена для зберігання, захисту, опрацювання, використання і поширення інформації про особу та документи, що оформлюються із застосуванням засобів Реєстру (національний та закордонний паспорти, посвідчення водія, картка мігранта, посвідчення біженця).

У реєстр вносяться такі особові данні: ім'я особи; дата народження/смерті; місце народження; стать; дата внесення інформації про особу до Реєстру; відомості про батьків (усиновлювачів), опікунів, піклувальників та інших представників; відомості про громадянство або його відсутність та підстави набуття громадянства України; реквізити документів, виданих особі (тип, назва документа, серія, номер, дата видачі та уповноважений суб'єкт, що видав документ, термін дії документа); відомості про документи, що підтверджують смерть особи або визнання особи померлою чи безвісно відсутньою. Також реєстр передбачає внесення біометричних даних — відцифрований зразок підпису особи, образ обличчя особи та відцифровані відбитки пальців рук. Додатково включається змінна інформація (про місце реєстрації, сімейний стан, про відмову від прийняття реєстраційного номера облікової картки платника податків, про видачу приватизаційних паперів).

На жаль, демографічний реєстр в Україні поки ще не функціонує в повному обсязі через відсутність належного програмного і технічного оснащення. До того ж він не відповідає своєму основному призначенню, оскільки має ознаки наглядово-фіскального характеру.

Загалом реєстр населення має використовуватися здебільшого для точного його обліку, в тому числі в комбінації із традиційним переписом чи вибірковими обстеженнями.

В Україні діють також спеціалізовані єдині державні реєстри. Наприклад, у фінансово-банківській системі — реєстри кредиторів, платіжних вимог, іпотек, реєстр цінних паперів тощо. Або у правовій сфері — реєстр судових рішень; реєстр осіб, які вчинили корупційні правопорушення, реєстр адвокатів та інші. Також ведуться державні реєстри в системі охорони здоров'я — реєстр лікарських засобів; реєстри хворих на: туберкульоз, цукровий діабет, канцер-реєстр тощо.

Матеріали статистичних реєстрів використовуються державними установами в рамках затверджених угод про взаємний обмін адміністративними даними. До слова, Держстат України має угоди про щоквартальний (щомісячний) обмін інформаційними ресурсами з такими державними установами України: Адміністрація спеціальної служби спеціального зв'язку та захисту інформації; Державна казначейська служба; Державна податкова служба; Державна митна служба; Державна міграційна служба; Державна служба зайнятості; Міністерство внутрішніх справ; Міністерство юстиції; Міністерство економічного розвитку і торгівлі; Міністерство екології та природних ресурсів; Пенсійний фонд; Національний банк; Фонд державного майна.

Сучасні статистичні реєстри в комбінації з адміністративними даними різних державних установ та відомств створюють потужні і різнопланові інформаційні системи. Результати статистичного спостереження сукупності одиниць за єдиним планом спостереження утворюють так звані *інформаційні масиви*. Своєю чергою, інформаційні масиви агрегуються, обробляються, аналізуються і у формі інформаційних потоків надходять до користувачів. *Інформаційний потік* — організований рух статистичних даних у межах інформаційної системи від джерел інформації до користувача різного рівня (Адміністрація президента, Верховна Рада, Кабінет міністрів, міністерства, відомства).

Потужним і здебільшого загальнодоступним джерелом узагальнених даних є статистичні банки даних. *Статистичний банк даних (СБД)* — електронна система інтегрованих інформаційних статистичних ресурсів та уніфікованих методичних матеріалів, так званих «мета-даних» (термінологічний словник, перелік використаної нормативно-законодавчої, наукової, методичної та довідникової літератури) з оперативним порядком доступу. У загальному вигляді СБД включає: автоматизований банк класифікаторів і класифікацій; електронний каталог статистичних показників; генеральну сукупність об'єктів статистичного спостереження; центральну та оперативну базу статистичних макро- і мікроданих, а також регіональні, галузеві і тематичні банки статистичних даних (рис. 3.4).

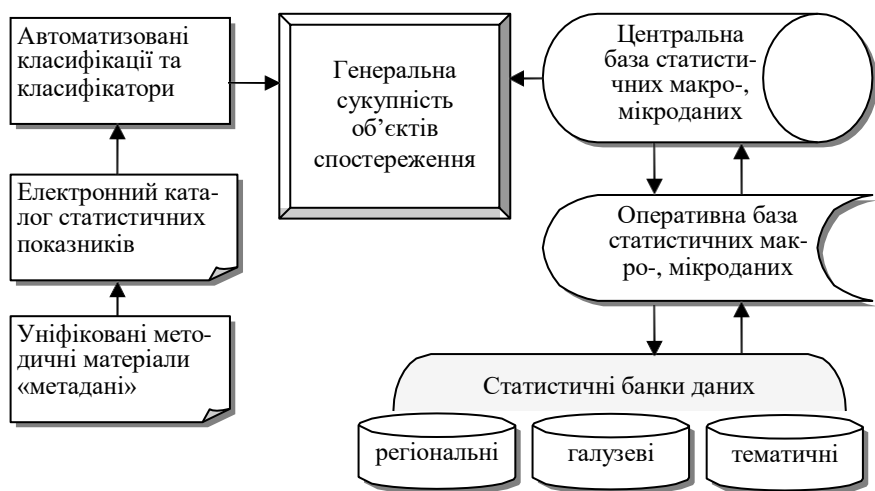


Рис. 3.4. Структура статистичних банків даних

Національні статистичні банки даних — статистичні таблиці, побудовані за ієрархічним принципом, що містять річну або кварталну архівну інформацію за тривалий (кількарічний) період. Використовуються також інші форми візуалізації та подання статистичної інформації, яка міститься у банках даних. Зокрема, на веб-порталі Держстату України розміщено статичні та динамічні

графіки: полігони, гістограми (у тому числі статево-вікові піраміди), секторні діаграми, а також картограми. Так, для зручного доступу до регіональної інформації створено інтерактивну картограму «Статистичний атлас регіонів України», що дозволяє ознайомитися з усіма показниками соціально-економічного розвитку загалом по Україні і окремо по кожній області та м. Києву. Користувачам також доступні інтерактивні тематичні карти України (наприклад, Ринок праці, Доходи та умови життя, Національні рахунки тощо). Статистичні банки даних, якими можна скористатися на сайті Держстату України, містять такі розділи: Комплексна статистика, Регіональна статистика, Економічна статистика (Макроекономічна статистика, Національні рахунки), Демографічна та соціальна статистика. Приміром, для управління соціально-демографічними процесами використовується банк даних України щодо чисельності і складу населення, його розміщення і процесів відтворення «Всеукраїнський перепис населення 2001», «Статистика населення України», де зберігаються архівні дані, від 1998 року.

Гармонізація електронних банків даних державних установ дає змогу в першу чергу оптимізувати систему документообігу і тим самим підвищити якість обліку завдяки автоматизованій процедурі узгодження адміністративних даних. Вони також вирішують проблему зайвого звітного навантаження на респондентів через зменшення багаторазових запитів різними інстанціями. Важлива перевага електронних банків даних — оперативний доступ до деперсоніфікованих даних та широке використання їх в аналітичних цілях. Прозорість інформаційних ресурсів (у рамках, передбачених законодавством) дозволяє громадськості відстежувати процеси, що відбуваються в суспільстві, і контролювати дії органів влади.

Інтеграція телекомунікаційної та комп'ютерної мереж в єдину систему сприяє оперативному, навіть «миттєвому» зв'язку між носіями інформації, органами її систематизації й опрацювання споживачами інформаційного продукту. Усе це допомагає ефективному прийняттю й упровадженню управлінських рішень.

Вправи та запитання для самоконтролю

1. Зазначте сфери застосування перепису як суцільного обстеження.
2. Чи може бути перепис несуцільним обстеженням і неперервним у часі?
3. Який спосіб статистичного спостереження дає найточніші дані про населення?
4. Чи використовуються дані несуцільних обстежень для визначення чисельності та складу населення?
5. Яким чином перепис населення комбінується з іншими видами і формами спостереження?
6. Які сучасні комбіновані моделі перепису населення є найбільш оперативними та економними?
7. Яку функцію щодо обліку економічних, соціальних і демографічних параметрів виконують державні статистичні реєстри?
8. Чи можуть статистичні реєстри замінити поточний облік?
9. Здійсніть порівняльний аналіз програми Всеукраїнського перепису населення України, проведеного у 2001 році, та проекту програми Всеукраїнського перепису населення України 2020 року. Скористайтеся відповідними матеріалами, розміщеними на сайті Держстату України.
10. Які новації містить проект програми Всеукраїнського перепису населення України 2020 року?
11. Чи містить переписний листок нового зразка запитання, які б допомогли встановити реальні розміри міграційного руху населення?
12. Чи передбачається проектом програми перепису визначення чисельності та статево-вікового складу внутрішньо переміщених осіб?
13. Розробіть проект програми загальнодержавного перепису сільськогосподарських тварин, які утримуються в аграрних господарствах (сільськогосподарські підприємства, фізичні особи-підприємці, домогосподарства, що займаються сільськогосподарською діяльністю), які загалом забезпечують 95 % валового виробництва продукції тваринництва.
14. Розробіть проект програми статистичного спостереження впроваджених українських стартапів у сфері дистанційного навчання (масові онлайн курси для студентів та школярів).
15. Розробіть програмно-методологічні питання плану (мета, об'єкт та одиниця спостереження, одиниця сукупності, перелік запитань, статистичний формуляр, форма контролю результатів) наведених далі статистичних спостережень:
 - складання електронного реєстру листків непрацездатності з урахуванням його номера, дати і тривалості відкриття, статусу медичного працівника тощо;

- єдиного реєстру отримувачів гуманітарної допомоги (вантажів, коштів в іноземній та національній валюті, наданих послуг, виконаних робіт);

- переліку впроваджених енергозберезувальних технологій в усіх адміністративно-територіальних одиницях з метою визначення: кількості та виду технологій; обсягів інвестицій, у тому числі іноземних; економічної ефективності з урахуванням сфери застосування.

16. Розробіть організаційні питання плану (місце, час, орган, персонал, вид, форма та спосіб спостереження) наведених далі статистичних спостережень:

- поточної перевірки суб'єктів господарювання з питань дотримання податкової дисципліни;
- складання картотеки клієнтів державної страхової компанії;
- реєстрації проданих квартир на первинному ринку житлової нерухомості.

17. Опишіть програмно-методологічні та організаційні питання плану розроблення статистичного реєстру підприємств, які мають заборгованість із виплати заробітної плати.

18. Розробіть програмно-методологічні та організаційні питання плану статистичного спостереження процесу децентралізації та самоврядування в країні. Чи використовується при цьому комбінація різних видів і способів спостереження? Якщо так, обґрунтуйте, які саме.

19. Опишіть програмно-методологічні та організаційні питання плану розроблення інформаційно-аналітичної бази трудових мігрантів. Чи використовується при цьому комбінація різних видів і способів спостереження? Якщо так, обґрунтуйте, які саме. В разі застосування статистичного реєстру опишіть його структурні елементи.

20. Опишіть структурні елементи (ідентифікаційні, стратифікаційні, реєстраційні і додаткові ознаки) наведених далі статистичних реєстрів:

- єдиного реєстру отримувачів житлових субсидій;
- єдиного державного реєстру транспортних засобів;
- єдиної державної електронної бази з питань освіти (ЄДЕБО).

21. Зазначте основні структурні елементи статистичних банків даних.

22. Опишіть характерні риси інформаційних ресурсів сучасних статистичних банків даних.

23. Визначте суть та практичне призначення уніфікованих метаданих.

24. Зазначте коло користувачів інформаційних потоків статистичних банків даних.

25. Що є індикаторами статистичного банку даних кон'юнктурних обстежень ринку товарів та послуг?

Рекомендована література

1. *Валенте П.* Проведение переписей в Европе: как считают население в раунде 2010 года // Вопросы статистики. — 2012. — № 12. — С. 3–8.
2. *Збарская И.А.* Перепись населения: вызовы и перспективы раунда 2020 года // Вопросы статистики. — 2012. — № 10. — С. 9–13.
3. *Єріна А.М., Пальян З.О.* Статистика: Підручник. — К. : КНЕУ, 2010. — 351 с.
4. *Пальян З.О.* Демографія : Навч. посіб. — К. : ВПЦ «Київський університет», 2014. — 222 с.
5. Перший Всеукраїнський перепис населення: історичні, методологічні, соціальні, економічні, етнічні аспекти / За наук. ред. І. Ф. Кураса, С.І. Пирожкова. — К. : ЮЦ Держстату України, 2004. — 558 с.
6. Рекомендации Конференции европейских статистиков по проведению переписей населения и жилищного фонда в 2020 году. — Нью-Йорк ; Женева : ООН, ECE/ces/2015/41. — 259 с.
7. *Стеценко С. Г.* Демографічна статистика : Підручник. — К. : Вища школа, 2005. — 415 с.

РОЗДІЛ 4

СТАТИСТИЧНІ МОНІТОРИНГИ

Статистичний моніторинг як неперервне спостереження
Первинні та сучасні форми моніторингу
Економічний та фінансовий моніторинг
Соціальний моніторинг: соціологічні і статистичні обстеження

Після вивчення цього розділу студент має

- *знати*: сутність статистичного моніторингу, його роль у статистичних дослідженнях і сфери застосування; особливості організації і проведення моніторингів: економічних, фінансових, екологічних і соціальних;
- *уміти*: розробляти план спостереження, програмно-методологічні та організаційні питання моніторингу як неперервного спостереження; здійснювати контроль якості отриманих даних; використовувати результати моніторингу в аналізі соціально-економічних явищ і процесів.

Статистичний моніторинг як неперервне спостереження

У структурі сучасних видів спеціально організованих статистичних обстежень моніторингам належать окрема ніша. Власне моніторинг означає постійне спостереження (стеження) за явищами і процесами. Тобто під час реєстрації характеристик явищ, об'єктів чи процесів постійно фіксується зміна їхнього стану чи перебігу.

Статистичний моніторинг — це спеціально організоване комплексне неперервне стеження за зміною стану певного середовища (явища, об'єкта) або перебігу процесу в часі чи просторі з регулярною реєстрацією сталого набору його кількісних та якісних параметрів.

У системі державної статистики використовується більш узагальнене тлумачення статистичного моніторингу як процесу спостереження за станом і зміною соціально-економічних явищ на підставі статистичної інформації.

У вузькому сенсі статистичний моніторинг — це систематичне стеження за зміною масових явищ і процесів. Утім, функціонування електронної (цифрової) економіки потребує комплексного

підходу до організації інформаційних ресурсів: отримання, опрацювання і використання. Тому в сучасних умовах застосовують ширше поняття статистичного моніторингу — регулярне збирання, систематизація, опрацювання, архівація, аналіз даних, прогноз та надання інформації користувачеві [Пугачева, 2007].

Виходячи з цього, статистичний моніторинг має виконувати такі завдання:

- реєстрація показників об'єкта статистичного спостереження за станом на окремі дати чи за інтервали часу;
- аналіз динаміки показників та прогнозування їхніх рівнів;
- імітаційне моделювання показників об'єкта спостереження, визначення ймовірних змін та оцінювання можливих ризиків;
- підгрунтя для оперативного або стратегічного управління об'єктом спостереження;
- забезпечення потреб користувачів у регулярних даних;
- інформаційне забезпечення інших видів спостереження (за сіб оновлення даних статистичних реєстрів, здійснення обліків) та створення і поповнення статистичних банків даних.

Беручи до уваги широту завдань та розмаїття об'єктів спостереження, статистичні моніторинги можна класифікувати за кількома аспектами: метою організації; часом проведення; рівнем узагальнення; специфікою даних (рис. 4.1).

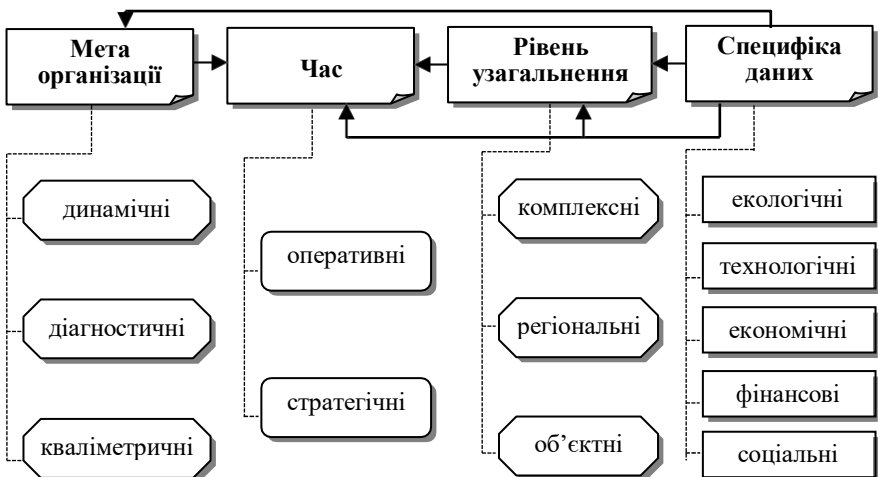


Рис. 4.1. Класифікація статистичних моніторингів

За метою організації статистичні моніторинги поділяються на динамічні, діагностичні (прогностичні) та кваліметричні.

Динамічний моніторинг — відстежування зміни параметрів об'єкта спостереження й аналіз інтенсивності і тенденції розвитку (коливання цін, заборгованості з виплат, перебігу військових дій, поширення епідемічних захворювань, зміни контенту ресурсів інтернету та ЗМІ тощо).

Діагностичний (прогностичний) моніторинг — це стеження за зміною параметрів об'єкта спостереження та виявлення причин і очікуваних змін (кліматичних умов, геофізичних параметрів довкілля, забруднення навколишнього середовища, попиту та пропозицій на ринку, економічних ризиків, демографічних загроз, соціально-політичних настроїв і подій).

Кваліметричний моніторинг — регулярний контроль якості об'єкта спостереження шляхом кількісного оцінювання його характеристик (якості продукції, товарів та послуг, результатів навчання, рівня кваліфікації, якості інформації тощо).

За часом проведення статистичні моніторинги можна розподілити на дві групи — оперативні та стратегічні.

Оперативний моніторинг — це щоденна, щотижнева, щомісячна, щоквартальна реєстрація даних (метеоумов, зареєстрованих заяв на послуги, дорожньо-транспортних пригод, результатів біржових торгів, обсягів зовнішніх міграційних потоків), якості (продукції, виконаних робіт), бюджетний моніторинг.

Стратегічний моніторинг — багаторічне спостереження за станом явищ і перебігом тривалих процесів (збереження окремих видів флори і фауни, викидів в атмосферу парникових газів, поширення небезпечних хвороб, соціально-економічного й демографічного розвитку територій, виконання міжнародних торгових зобов'язань).

За рівнем узагальнення вирізняють моніторинги: комплексні, регіональні, об'єктні.

Комплексний моніторинг охоплює одночасно кілька об'єктів спостереження. Прикладами такого моніторингу можуть бути спостереження за зміною: кліматичних умов; екологічних, демографічних та економічних наслідків техногенних катастроф. В економічній сфері на загальноєвропейському рівні здійснюється моніторинг економічного розвитку країн-членів ЄС та характеру економічних перетворень у країнах, що прагнуть вступити до Євросоюзу. На національному рівні ведуться моніторинги: втілення

економічних реформ; боротьби з корупцією в органах управління та правозахисній системі; подолання бідності населення. У соціальній сфері на міжнародному рівні об'єктами статистичного моніторингу є: гендерна рівність, гідна праця, забезпечення освітио упродовж життя тощо. У системі Держстату України здійснюється моніторинг якості статистичної інформації та її відповідності європейським і міжнародним стандартам.

Регіональний моніторинг стосується спостереження за явищами і процесами, що відбуваються на національному, локальному або місцевому рівні. Скажімо, загальнонаціональними є моніторинги втілення реформ у правоохоронній та судовій системі, у сфері охорони здоров'я. У соціально-економічній сфері — моніторинг процесів регіональної децентралізації та місцевого самоврядування. В екології це моніторинг рівня радіаційного забруднення на територіях, що постраждали внаслідок аварії на ЧАЕС, або моніторинг рівня підйому води у басейнах річок на період весняних паводків чи вирубки лісів у гірських районах місцевості.

Об'єктний моніторинг зазвичай стосується конкретного явища чи процесу.

Наприклад, контроль за рівнем цін на місцевому ринку нафтопродуктів або за державними закупівлями сільгосппродуктів; моніторинг рівня забруднення водойм регіону стічними водами чи заборгованості з виплати заробітної плати та соціальної допомоги, або моніторинг обсягів біржових торгів, вартості укладених угод та котирування цін на біржах.

За *специфікою даних* статистичні моніторинги можна поділити на екологічні, технологічні, фінансові та економічні і соціальні.

Екологічні моніторинги відстежують явища і процеси у природному середовищі, ґрунтуються на результатах метеорологічних, атмосферних, гідрологічних, геофізичних і геодезичних, агротехнічних, радіологічних, аерокосмічних спостережень. Такі моніторинги мають на меті не лише фіксувати зміни у станах відповідних об'єктів спостереження з урахуванням факторів впливу, а й прогнозувати і моделювати ймовірні зміни, оцінювати їхні можливі наслідки. Об'єктами спостереження можуть бути як кліматичні умови, так і стан довкілля в контексті господарської діяльності людей: ефективність використання природних ресурсів, запасів корисних копалин, земельного і лісового фонду, водних ресурсів; рівень забруднення довкілля й відшкодування наслідків забруднення; утворення та утилізація промислових і

побутових відходів. До цієї групи належить агрометеорологічний моніторинг, що за сучасних умов здійснюється шляхом аерокосмічного спостереження (супутникового зондування і зйомки). Завданням цього моніторингу є оцінювання якості ґрунтів та запасу вологи у ґрунтах під сільськогосподарськими угіддями, стану посівів та їхньої біологічної продуктивності, раціональної сівозміни та визначення очікуваного врожаю, а також реальних темпів збирання врожаю та його обсягу. Окремий напрям екологічного моніторингу — оцінювання впливу негативних екологічних чинників на стан здоров'я населення.

Технологічні моніторинги мають на меті контролювати технічний стан об'єктів у процесі їхньої експлуатації, діагностувати аварійні загрози, контролювати якість роботи об'єктів та формувати банк поточних і архівних результатів контролю. Різновидом технологічних моніторингів є спостереження за якістю продукції, дотримання технологічних норм і стандартів щодо її виробництва, а також моніторинг якості товарів та послуг (контроль якості й безпечності харчових продуктів; якості, повноти і своєчасності надання послуг тощо).

Економічний моніторинг здійснюється шляхом регулярної реєстрації показників економічного розвитку (країни чи її регіонів; видів економічної діяльності, суб'єктів господарювання). Скажімо, бюджетний моніторинг, який щоквартально здійснює Державна казначейська служба спільно із Держстатом України, має на меті відстежувати ефективність виконання бюджетних програм і цільове використання коштів державного і місцевих бюджетів. Або інші приклади: моніторинг біржових торгів (валютних, товарних, цінних паперів) чи коливання цін на внутрішньому та зовнішньому ринках, зокрема цін на пальне, зерно, олію, цукор тощо.

Соціальний моніторинг покликаний здійснювати регулярні спостереження за явищами і процесами, що відбуваються в суспільному житті. Окрім реєстрації відповідних статистичних показників цей моніторинг фіксує також результати соціологічних обстежень населення.

Для проведення статистичного моніторингу складається план спостереження, тобто розробляються його програмно-методологічні та організаційні питання. Мета спостереження визначає вид моніторингу та його конкретні завдання. Наприклад, моніторинг кон'юнктури ринку товарів та послуг проводиться з метою оцінювання задоволення попиту населення і

сприяння ефективному збуту виробленої продукції. Встановлюється об'єкт спостереження й обґрунтовуються одиниці сукупності та одиниці спостереження з урахуванням цензу (ціни на товари-представники або соціально значущі товари). Складається програма спостереження, а саме обґрунтується система індикаторів та показників, які реєструватимуться на постійній основі.

Слід зауважити, що під час проведення моніторингу об'єктивний та суб'єктивний час спостереження зазвичай збігається, оскільки реєстрація даних про явища і процеси здійснюється мірою їхнього виникнення. На етапі розроблення програми спостереження обґрунтовується процедура гармонізації результатів моніторингу з іншими видами та формами спостереження.

У програмі моніторингу описується порядок контролю якості даних. У рамках логічного контролю практикується зіставлення результатів моніторингу з аналогічними даними інших форм і способів спостережень. Під час арифметичного контролю вдаються, наприклад, до агрегування даних та порівняння їх з відповідними макропоказниками або аналізується характер розсіювання даних та наявні екстремальні «викиди». Якщо моніторинг спирається на дані вибіркового обстеження, розраховуються похибки репрезентативності.

Окремий пункт програми моніторингу — це визначення виду і способу спостереження. Оскільки більшість сучасних моніторингів мають комплексний характер, то й спостереження може виконуватися шляхом комбінації даних звітності, обліків, а також із залученням матеріалів переписів та вибіркового спостереження, статистичних реєстрів і адміністративних даних. Залежно від виду і способу спостереження розробляють статистичний інструментарій (бланки звітності, опитувальні листки, електронні таблиці тощо).

На етапі розроблення програми спостереження визначають коло споживачів результатів моніторингу.

Первинні та сучасні форми моніторингу

Уперше моніторинг як послідовність регулярних спостережень почали застосовувати в екології. Здійснювався він для відстежування в часі стану довкілля, оцінювання впливу факторів та прогнозування ймовірних змін. Зокрема, об'єктом моніторингу

ставали метеорологічні умови, стан ґрунтів, ґрунтових вод, лісових ділянок або рівень забруднення довкілля (атмосферного повітря, водних басейнів, ґрунтів, радіаційного забруднення).

До первинних форм належить моніторинг мінерально-сировинної бази в рамках глобальних, національних і локальних обстежень. Тобто йдеться про систему спостережень, збирання, опрацювання, передавання, зберігання та аналізу інформації щодо розвіданих запасів та ресурсів корисних копалин, прогнозування їхнього стану, потужності, видобутку та переробки. Результати такого моніторингу в Україні використовують для формування Державного фонду корисних копалин та його резервів.

Сучасні світові й регіональні екологічні та агротехнічні проблеми диктують появу нових об'єктів і форм моніторингу.

Глобальні моніторингові обстеження ведуться з метою визначення антропогенного впливу екологічного забруднення. Актуальним є отримання даних щодо надходження забруднювальних речовин та парникових газів у атмосферне повітря. Україна також здійснює моніторинг у рамках виконання Кіотського протоколу та Паризької угоди (від 12.12.2015 року) про зменшення викидів парникових газів. Статистичний моніторинг виконується шляхом ведення реєстру викидів парникових газів, а також подання щоквартальної та річної статистичної звітності. Звітність організується як несуцільне спостереження (основного масиву) з урахуванням міжнародних рекомендацій Статистичної комісії ООН щодо інвентаризації атмосферних викидів (CORINAIR'2000).

Окремо здійснюється моніторинг акваторій — багаторівневий моніторинг середовища, який включає дослідження на дні акваторії за допомогою донних станцій, космічний моніторинг, контактні вимірювання (проб води у зоні нафтовидобутку). Мета — виявлення нафтових забруднень і вимірювання товщини нафтових плівок на поверхні моря, океану).

З метою налагодження ефективного використання земельного фонду здійснюється статистичний моніторинг землекористування (сільськогосподарських угідь, земель запасу, зрошувальних, осушених, зволжених земель), а також площі забруднених земель, що містять радіоактивні компоненти, важкі метали, шкідливі хімікати, біологічні отруйні речовини. Для визначення екологічного стану територій, вкритих лісами, та збереження лісів здійснюється моніторинг лісокористування (санітарна чистка, відновлення, вирубка), а також площі лісів, які постраждали вна-

слідок пожеж, ураження шкідниками та через незаконну вирубку. Організується статистичний моніторинг у формі суцільної статистичної звітності.

У теперішній час моніторинг вийшов за межі суто природних явищ та процесів. Його використовують в економічній, соціальній та політичній сферах. Водночас розширюються й завдання моніторингу. Сьогодні це не лише інформаційна система постійної реєстрації та оновлення даних про об'єкт, а й відповідна технологія оцінювання поточного стану об'єкта, його прогнозування та прийняття адекватних управлінських рішень.

В економічній, зокрема у фінансовій сфері провадиться моніторинг інвестиційного кредитування, завданням якого є аналіз балансу та фінансового становища позичальників, контроль за виконанням умов угоди, виявлення проблемних інвестиційних кредитів. З боку інвесторів відбувається спостереження за діями керівництва компанії, наприклад, за обслуговуванням заборгованості й дотриманням термінів її погашення. Такий моніторинг організується щомісяця або кожні три місяці.

В Україні здійснюється моніторинг ринку праці, а саме зайнятості, безробіття та працевлаштування. Інший приклад — моніторинг бюджетів окремих соціальних груп населення (пенсіонерів, молодих сімей, однаків).

У соціальній сфері запроваджено статистичний моніторинг соціальних, у тому числі електоральних настроїв населення, моніторинг якості освіти випускників середніх навчальних закладів та якості підготовки фахівців закладами вищої освіти. Прикладом політичного моніторингу є так званий Громадський моніторинг виконання Україною плану дій «Україна — ЄС».

Стрімкий розвиток інформаційних і телекомунікативних технологій зумовив появу нових моніторингових систем. Загальновідома *Система глобального позиціонування (Global Positioning System)* — визначення положення та швидкості руху об'єкта на поверхні земної кулі або в атмосфері за допомогою спеціальних радіоелектронних засобів. Положення об'єкта обчислюється за допомогою GPS-приймача, який приймає й обробляє сигнали супутників космічного сегмента GPS. Для визначення точних параметрів орбіт супутників та керування GPS-система має наземні центри управління. GPS-система широко використовується у навігації (наземній, атмосферній та водній). Тепер вже важко уявити рух транспорту на автошляхах та сучасних автострадах без

GPS-навігатора. За допомогою GPS-смартфонів фіксуються затори й ДТП на дорогах, дорожні ремонтні роботи, якість доріг (яма і нерівності) і таким чином прокладається оптимальний маршрут руху автомобілів.

Ще однією надзвичайно популярною сучасною формою спостереження є *моніторинг Інтернет- та медіаресурсів*. Моніторинг інтернет-ресурсів — це спостереження за змістом і зміною тем, подій та інформації, розміщених в інтернеті. Він включає пошук, аналіз, систематизацію й класифікацію інформації за тематикою, яка цікавить користувача. Спостереження здійснюється як шляхом самостійного перегляду користувачами кількох веб-сайтів за обраною тематикою, так і за допомогою пошукових систем і каталогів.

Запровадження сучасних моніторингових систем значно інтенсифікує процес відстеження та пошуку інформації в інтернеті. Безперечною перевагою моніторингових систем є їхня оперативність, оскільки бази даних поповнюються кожні 10–20 хвилин, а джерела скануються з інтернету мірою їх оновлення. На відміну

від пошукових систем моніторингові мають архівний банк, де зберігається видалена з веб-сайта інформація. Завдяки аналітичному інструментарію користувачі можуть не лише отримувати інформацію в режимі реального часу, а й власноруч формувати дайджести, аналізувати інформаційні зміни, систематизувати схожі за змістом новини, розміщені на різних веб-сайтах, тощо.

Розмаїття сучасних сфер застосування статистичного моніторингу потребує їхнього диференційованого розгляду.

Економічний та фінансовий моніторинг

Економічний моніторинг — система взаємопов'язаних регулярних спостережень щодо зміни показників економічного розвитку. Залежно від рівня узагальнення об'єкта спостереження економічний моніторинг може проводитися на міжнародному, національному чи локальному рівнях.

У міжнародній практиці використовують моніторинги різних процесів, що притаманні ринковій економіці. Необхідність таких обстежень диктується умовами конкуренції на внутрішньому і зовнішньому ринках. Ефективне управління й регулювання економічних процесів неможливе без систематичної реєстрації та

аналізу відповідних характеристик їхнього розвитку. До числа моніторингів, здійснюваних за міжнародними стандартами, належать моніторинги: ділової активності підприємств, кон'юнктури ринку товарів та послуг, економічної активності населення, доходів і витрат домогосподарств тощо.

Статистичний моніторинг ділової активності підприємств — статистичне спостереження керівників підприємств (фірм) стосовно поточного стану та перспектив розвитку підприємств. Він здійснюється з метою якісного аналізу поточних умов та чинників розвитку підприємств, побудови короткострокових прогнозів, забезпечення додаткової інформації для системи національних рахунків. Моніторинг ділової активності підприємств належить до кон'юнктурних обстежень і має певні переваги:

- оперативність (одержання даних упродовж 15–20 днів);
- прогностичність (оцінюється не лише поточний стан, а й очікувані зміни);
- інформативність (одержання даних про види, сфери й аспекти діяльності, недостатньо охоплювані звітністю).

Завданнями обстеження ділової активності підприємств є:

- оцінювання поточної ситуації на внутрішньому ринку товарів та послуг;
- визначення найближчих намірів діяльності підприємств та кон'юнктурної поведінки;
- оцінювання можливих ризиків і загроз функціонуванню підприємств;
- визначення короткострокових перспектив розвитку виду діяльності, галузевої структури регіону;
- уточнення макроекономічних показників;
- інформування агентів ринку про стан і перспективи розвитку внутрішнього ринку та потрапляння на зовнішній ринок;
- відстежування періодів зростання та спаду економіки з місячною або кварталною періодичністю (визначення фази бізнес-циклів: підйом, рецесія, депресія, пожвавлення).

Моніторинг фіксує індикатори ділових очікувань, які розраховують в Україні відповідно до гармонізованої програми ЄС для підприємств (фірм) таких видів економічної діяльності: промисловість, будівництво, роздрібна торгівля, сфера послуг, у тому числі фінансових. До складу індикатора ділових очікувань входять субіндикатори: ділової впевненості; ділового клімату та індикатор економічних настроїв.

Джерелом інформації для розрахунку індикаторів ділових очікувань в Україні є дані щоквартального державного статистичного спостереження «Стан ділової активності підприємств» за формами, які запроваджені у статистичну діяльність від 2015 року. Зокрема, розрахунок індикатора ділового клімату здійснюється в динаміці з I кварталу 2015 року, а індикатор економічних настроїв у динаміці з I кварталу 2017 року.

Інформаційну базу формують відповіді керівників на запитання анкет, адаптованих для окремих видів діяльності. Так, «Анкета для промислового підприємства» (форма № 2К-П квартальна) містить такі показники ділової впевненості: оцінка поточного обсягу замовлень на виробництво та запасів готової продукції, очікувані зміни в обсязі виробництва впродовж наступних трьох місяців. Показниками ділового клімату є: зміни в обсязі виробництва впродовж минулих трьох місяців та очікувані зміни протягом наступних трьох місяців; оцінка поточного обсягу замовлень на виробництво продукції, у тому числі обсягу іноземних замовлень.

Для узагальнення й інтерпретації відповідей керівників підприємств по кожному з показників розраховується різниця (баланс) між часткою респондентів, які надали взаємозаперечувальні відповіді («збільшиться» і «зменшиться»). Додатний або від'ємний баланс визначає напрям очікуваної тенденції зміни кожного показника.

Наприклад, відповіді на запитання щодо ділової впевненості керівників промислових підприємств записуються такі: поточний обсяг замовлень (попит) на виробництво продукції («+» більш ніж достатній або вище норми; «=» достатній або нормальний для сезону; «-» недостатній або нижче норми). Оцінювання поточних запасів готової продукції здійснюється у такій спосіб: «+» дуже великий (вище норми); «=» достатній (нормальний для сезону); «-» дуже малий (нижче норми). На запитання: «Як, на Вашу думку, зміниться обсяг виробництва продукції в наступні три місяці?», дається одна з трьох відповідей: «+» збільшиться; «=» не зміниться; «-» зменшиться.

На підставі балансів відповідей обчислюється індикатор ділової впевненості у промисловості станом на момент t як середня арифметична проста із сезонно скоригованих значень балансу трьох оцінок: поточного обсягу замовлень (попиту) на виробництво продукції x_1 , обсягу запасів готової продукції x_2 , зміни обсягу

виробництва продукції в наступні три місяці хз. Сезонне коригування здійснюється методом згладжування квартальних даних за попередні чотири роки.

Індикатор ділового клімату (у промисловості) обчислюється методом головних компонент на підставі сезонно скоригованих значень:

- балансів обсягу виробництва продукції за минулі та у наступні три місяці;
- оцінок поточного обсягу замовлень (попиту) на виробництво продукції;
- поточного обсягу іноземних замовлень (експортного попиту) на виробництво продукції;
- поточного обсягу запасів готової продукції.

Індикатори ділової впевненості та ділового клімату порівнюють із сезонно скоригованими темпами приросту ВВП, а індикатор економічних настроїв — із сезонно скоригованими темпами зростання ВВП. На підставі індикатора ділового клімату будують графік дослідження «Бізнес-циклів».

Моніторинг кон'юнктури ринку товарів та послуг — це систематична реєстрація статистичних даних про стан ринку товарів та послуг, а також про підприємницький клімат з метою прийняття рішень щодо задоволення попиту населення й ефективного збуту виробленої продукції. Він також належить до числа кон'юнктурних спостережень, завданнями якого є:

- надати об'єктивну інформацію про вітчизняних товаровиробників та внутрішніх конкурентів, оцінити ситуацію на ринку та спрогнозувати її розвиток;
- досягти конкурентних переваг і знизити рівень ризику банкрутства (пошук вільного та вигідного сегмента ринку);
- визначити правильну стратегію диверсифікації (встановити оптимальний рівень цін на певні товари та послуги).

У рамках кон'юнктурних спостережень ринки класифікують за різними ознаками: рівнем організації ринкової інфраструктури; функціональним призначенням об'єкта продажу; територіальною ознакою; обсягом продажу товарів, масштабом і рівнем конкуренції.

За рівнем організації ринкової інфраструктури ринки поділяються на: організовані (біржова, страхова і банківська діяльність, товарні ярмарки, науково-технічні послуги тощо) та неорганізовані — функціонують переважно за принципом прямих

зв'язків контрагентів ринку. Особливістю організованих ринків з точки зору моніторингу кон'юнктури є наявність інформації про характер попиту та пропозиції. Наприклад, на товарно-сировинних і фондових біржах інформація повніша порівняно із позабіржовим ринком завдяки реєстрації біржових угод. Біржі публікують інформацію про торги й забезпечують певний аналіз кон'юнктури на підставі відповідних індексів. Неорганізовані ринки меншою мірою схильні до різних коливань кон'юнктури, проте розкид показників кон'юнктури в них більший, ніж на організованих.

За функціональним призначенням об'єкта продажу ринки поділяють на: товарний, ринок нерухомості, землі, праці, капіталу (кредитно-грошовий, ринок інвестицій, ринок цінних паперів, валютний), послуг.

За територіальною ознакою виокремлюють ринки: внутрішній (регіональний), національний, світовий.

Згідно із методикою Міжнародного валютного фонду виокремлюють такі групи послуг: транспортне обслуговування (пасажирські та вантажні перевезення); туристичні послуги; зв'язок; будівництво; страхування; фінансові; комп'ютерні й інформаційні послуги; роялті та ліцензування; інші бізнес-послуги (посередницькі, технічні, лізингові); персональні, культурні та рекреаційні послуги; послуги державних установ.

Моніторинг попиту та пропозиції на внутрішньому ринку товарів та послуг враховує динаміку цін, рівня доходів населення та обсягів виробництва, а також динаміку обсягів і структуру імпорту та експорту, зміну умов зовнішньої торгівлі та зміну кредитно-грошової і валютної ситуації.

Індикаторами кон'юнктури товарів та послуг є:

- стійкість, циклічність і динаміка розвитку ринку (середній рівень показників, їхня варіація, сезонні коливання, наявність тренду);

- ринкова пропозиція (обсяг, структура і динаміка продажу, еластичність товарів);

- ринковий попит (обсяг, структура і динаміка; сегментація і диференціація попиту; ємність ринку);

- пропорційність ринку (збалансованість попиту та пропозиції, структури споживання і товарообігу; поділ ринку між основними виробниками, торговими посередниками і торговельними закладами);

- ділова активність (обсяг, кількість, частота і динаміка укладання торговельних угод; портфель замовлень, його склад і динаміка); рівень комерційного ризику (інвестиційного, прийняття рішень, кон'юнктурних коливань).

Моніторинг динаміки цін на споживчому ринку здійснюється Держстатом України в рамках статистичного спостереження за змінами цін (тарифів) на споживчі товари (послуги) і розрахунку індексів споживчих цін від 2013 року. Територіальну мережу спостереження утворюють всі обласні центри включно із містами: Київ, Біла Церква, Маріуполь, Сєвєродонецьк.

Перелік продуктів харчування, що складається із 23 найменувань, формується за трьома критеріями: наявність попиту населення на продукти; регулярна наявність продуктів на споживчому ринку; набір продуктів має представляти споживчий ринок міського поселення. Окрім цін на продукти харчування моніторинг фіксує ціни на два види пального (бензин А-95 та дизельне пальне).

Паралельно із Держстатом України моніторинг цін на продукти харчування здійснює Міністерство економічного розвитку в рамках Проекту «Платформа ефективного регулювання». Причому відстежується не лише зміна цін на основні соціальні продукти (23 найменування), а й визначається собівартість страв підстановкою найменування інгредієнтів та їхньої ваги. Окрім того, реєструється щоденна динаміка цін на товари за даними найбільших онлайн-супермаркетів. Результати моніторингу розміщено на сайті, де агреговано дані Держстату України, Нацбанку України, Світового банку та Євростату.

Важливою складовою моніторингу кон'юнктури ринку товарів та послуг є статистична звітність. Так, звітність колективних підприємств сільського господарства, фермерських та особистих підсобних господарств містить дані про обсяги виробленої та реалізованої продукції, а також про обсяги заготівлі рослинної і тваринної продукції. Статистична звітність підприємств торгівлі та послуг дає інформацію про обсяг і структуру роздрібного товарообігу.

Додатковим інформаційним ресурсом моніторингу споживчих цін є дані про структуру та рівень споживання товарів та послуг різних соціальних груп населення, що одержують в результаті вибіркового обстеження умов життя домогосподарств. Вибіркою охоплюються усі регіони України, за винятком АР Крим, невід-

контрольних Україні територій Донбасу, територій, що постраждали внаслідок аварії на ЧАЕС. Загальний обсяг вибірки — 13029 домогосподарств України, що поділяються за типом поселення на міські та сільські. Серед міських виокремлюють «великі міста» (100 тис. осіб і більше) та «малі міста» (до 100 тис. осіб). Обсяг вибірки на контрольованій частині території областей, охоплених операцією об'єднаних сил, становить відповідно: 866 домогосподарств в Донецькій області і 584 — в Луганській. Територіальна вибіркова мережа використовується без змін упродовж 5 років, а мережа домогосподарств — 1 рік.

Для проведення вибірових обстежень умов життя домогосподарств використовується такий статистичний інструментарій:

- контрольна картка домогосподарства;
- анкета основного інтерв'ю;
- щоденник поточних витрат;
- щоквартальний опитувальник щодо доходів і витрат;
- журнал поквартальних витрат.

Результати обстеження містять дані про місячне середньодушове споживання окремих продуктів харчування (груп в умовному перерахунку) в натуральних одиницях (кг, шт.) у розрізі: виду поселення (великі, малі міста, сільська місцевість); регіонів та економічних районів; типу домогосподарств (з дітьми, без дітей).

У фінансово-банківській сфері застосовується статистичний моніторинг діяльності кредитних установ. Специфіка цього виду діяльності зумовлює певні методологічні й організаційні особливості проведення спостереження. По-перше, орган спостереження, яким є Національний банк України, є водночас однією з одиниць спостереження. По-друге, моніторинг ґрунтується на комбінації різних форм і способів спостереження (адміністративні дані, звітність, опитування, реєстр). Причому переважна більшість статистичної інформації — адміністративні дані НБУ та Національної комісії з регулювання фінансових послуг (НКРФП). По-третє, жорстко дотримуються норми Регламенту Європейського парламенту, Ради та Комісії ЄС щодо організації структурної статистики.

Статистичним моніторингом охоплюються такі кредитні установи: Національний банк України; державні (Укрексімбанк, Ощадбанк, Укргазбанк, Приватбанк) та комерційні банківські установи, а також небанківські кредитні установи. Небанківські кредитні установи — це різноманітні фінансові посередники, які

акумулюють тимчасово вільні грошові кошти і розміщують їх у формі кредиту, цінних паперів, полісів, боргових зобов'язань тощо. До цих установ належать: страхові компанії, пенсійні фонди, ломбарди, лізингові та факторингові компанії, інвестиційні фонди, кредитні кооперативи.

Страхові компанії (страхування життя і майна) утворюються власними коштами засновників та продажем страхових полісів з подальшим розміщенням коштів в активах. З доходів від активів формується фонд відшкодування.

Пенсійні фонди (державні і приватні) формуються шляхом нарахувань на заробітну плату працівників або ж відрахувань з відповідних бюджетів, а розміщення коштів здійснюється переважно в державних цінних паперах (державні фонди) або в інших активах (приватні фонди).

Ломбарди видають позики населенню під заставу рухомого майна. Кошти утворюються за рахунок внесків засновників, прибутку від їхньої діяльності, виручки від реалізації заставного майна.

Діяльність лізингових компаній полягає в особливій, товарній формі кредитування (придбання та передавання в оренду товарів тривалого використання із поступовим погашенням вартості товару). Факторингові компанії здійснюють купівлю у фірм права на вимогу боргу у вигляді дебіторських розрахунків за поставлені товари, виконані роботи, надані послуги.

Інвестиційні фонди (банки, компанії) управляють вільними грошовими коштами інвестиційного призначення у формі цінних паперів. Кошти формуються із дивідендів і процентів за цінними паперами та з курсової різниці паперів. Окремою групою небанківських кредитних установ є кредитні кооперативи (товариства, спілки), які задовольняють потреби у кредитах членів кооперативів (підприємств малого і середнього бізнесу, фермерських та домашніх господарств, фізичних осіб). Кошти формуються продажем паїв своїм членам, стягуванням з них спеціальних внесків, одержанням позик у банку, одержанням доходів від поточної діяльності.

Для формування статистичного банку даних Держстат України використовує адміністративні дані, що надходять від НБУ (про діяльність НБУ та банків) та від НКРФП — щодо небанківських кредитних установ. Паралельно залучаються дані державних статистичних спостережень у рамках структурної статистики, а також Реєстр статистичних одиниць.

Мета статистичного моніторингу діяльності кредитних установ — отримання статистичної інформації про діяльність кредитних установ (банківських та небанківських), яка використовується для:

- складання показників структурної статистики щодо діяльності центрального банку та кредитних установ, що класифікуються за класом 64.19 «Інші види грошового посередництва» за КВЕД;

- складання показників структурної статистики кредитних установ, що класифікуються за класами 64.19 «Інші види грошового посередництва» та 64.92 «Інші види кредитування»;

- аналізу тенденцій розвитку кредитних установ забезпечення інформаційних потреб національних користувачів і міжнародних організацій;

- здійснення відповідних міжнародних порівнянь.

За мірою охоплення одиниць це суцільне обстеження, а за часом реєстрації даних — поточне із річною періодичністю. По кожному виду статистичних одиниць відстежується певна система статистичних показників. Так, моніторинг діяльності НБУ реєструє такі показники:

- витрати на персонал, у тому числі витрати на оплату праці та відрахування на соціальні заходи;

- кількість зайнятих і найманих працівників (у еквіваленті повної зайнятості);

- капітальні інвестиції за напрямками;

- доходи та витрати за їх видами (процентні доходи, комісійні доходи, дохід за рахунок акцій та інших цінних паперів, інші операційні доходи);

- чистий прибуток або чистий збиток за фінансовими операціями.

Статистичний моніторинг діяльності небанківських кредитних установ дає інформацію про: кількість підприємств (за правовим статусом; місцем знаходження материнської компанії; категорією кредитних установ) та кількість місцевих одиниць (за регіонами); чисельність зайнятих та найманих працівників (за категорією кредитних установ: установи, що спеціалізуються на наданні кредитів та інші кредитні установи); розмір капітальних інвестицій за напрямками; доходи та витрати за видами; чистий прибуток або чистий збиток за фінансовими операціями; суми кредитів та авансів клієнтам; розмір капіталу та наявних резервів. Результати моніторингу оприлюднюються Держстатом України, зокрема на

його веб-сайті у розділі «Публікації» / «Економічна статистика» / «Економічна діяльність» / «Діяльність підприємств».

Останнім часом в Україні запроваджено специфічний вид статистичного моніторингу — *фінансовий моніторинг*, який здійснюється для запобігання та протидії легалізації (відмиванню) доходів, одержаних злочинним шляхом, фінансуванню тероризму та фінансуванню розповсюдження зброї масового знищення. Національний банк України постановою Правління від 15 вересня 2016 року № 388 затвердив Положення про здійснення небанківськими фінансовими установами, які є платіжними організаціями та/або членами чи учасниками платіжних систем, фінансового моніторингу наданих ними фінансових послуг з переказу коштів.

Органами державного фінансового моніторингу є: Національний банк України, Міністерство фінансів України, Міністерство юстиції України, Міністерство транспорту та зв'язку України, Міністерство економіки України, Національна комісія з цінних паперів та фондового ринку, Національна комісія з регулювання фінансових послуг України, Державна служба фінансового моніторингу України.

Первинний фінансовий моніторинг охоплює широке коло одиниць спостереження: банки і згадані вище небанківські кредитні установи; біржі, платіжні організації та члени платіжних систем; компанії з управління активами.

Окрім того, одиницями спостереження є:

- оператори поштового зв'язку та інші установи, які проводять фінансові операції з переказу коштів;
- суб'єкти підприємницької діяльності, які надають посередницькі послуги при здійсненні операцій купівлі-продажу нерухомого майна;
- суб'єкти господарювання, які здійснюють торгівлю за готівку дорогоцінними металами та дорогоцінним камінням і виробами з них;
- суб'єкти господарювання, які проводять лотереї і азартні ігри, включно з казино, електронним казино;
- нотаріуси, адвокати, аудитори, аудиторські фірми, фізичні особи — підприємці, які надають послуги з бухгалтерського обліку, суб'єкти господарювання, які надають юридичні послуги.

Об'єктом фінансового моніторингу є фінансова операція, сума якої дорівнює чи перевищує 150 000 гривень (для суб'єктів гос-

подарювання, які проводять азартні ігри — 13 000 гривень) або дорівнює чи перевищує суму в іноземній валюті, еквівалентній 150 000 гривень (для суб'єктів господарювання, які проводять азартні ігри — 13 000 гривень). У згаданому Положенні НБУ чітко визначено, які за характером фінансові операції підпадають під контроль фінансового моніторингу.

У біржовій діяльності також здійснюється статистичний моніторинг. Відстежуючи дані про діяльність бірж, отримують статистичну інформацію про структуру пропозицій та угод, укладених на біржі, обсяги угод у натуральному та вартісному вираженні, структуру укладених угод за типом секцій і котирування цін на біржах загалом по країні та по кожному регіону окремо. Така інформація подається в розрізі видів діяльності бірж (товарні, фондові, валютні), їхньої форми власності (державні, приватні, змішані), а також форми участі в торгах: відкриті (члени біржі і охочі, які придбали право участі в торгах); закриті (лише члени бірж). Моніторинг біржової діяльності організується у формі статистичної звітності як суцільне поточне (щоквартальне) обстеження. Одиницею спостереження є біржа — юридична особа, яка здійснює біржову діяльність. Органом спостереження виступає Держстат України.

Первинна реєстрація даних біржових торгів здійснюється у формі статистичного моніторингу впродовж торговельної сесії (тижня, місяця, кварталу). Моніторингом реєструються, у першу чергу, показники цін біржових угод. Зокрема, фіксуються: аукціонні ціни, що формуються в результаті публічного змагання продавців і покупців; ціни первинного ринку (ціна відсікання і ціна погашення); ціни вторинного ринку (ціна виконання, відкриття, закриття, попиту та пропозиції).

Для характеристики динаміки та коливань цін біржових цінних паперів використовують статистичні показники:

- варіації для аналізу ризикованості вкладень у цінні папери;
- зміну ціни як різницю між ціною закриття торгової сесії поточного дня і ціною закриття торгової сесії попереднього дня;
- середньозважену ціну, тобто середню арифметичну зважену з цін, за якими виконано угоди з цінним папером протягом торгової сесії; вагами при цьому є кількість проданих цінних паперів;
- максимальну (мінімальну) ціну — це найвища (найнижча) ціна, яка зареєстрована в біржових угодах упродовж торгової сесії (тижня, місяця, кварталу). У міжнародній практиці часто ви-

користовують показники максимальної та мінімальної ціни, зареєстрованої протягом останніх 52 тижнів (останнього року).

Також моніторинг дає первинну й узагальнену інформацію для відстежування: обсягів і динаміки пропозицій та укладених угод; ліквідності та структури укладених угод; статутного капіталу бірж; кількості членів бірж і середньооблікової чисельності працівників, фонду оплати їхньої праці.

Для цього використовується статистична форма № 1-біржа (квартальна), що складається із чотирьох розділів. Показники перших трьох розділів заповнюються наростаючим підсумком з початку року, четвертий розділ складається за кожен квартал окремо. До першого розділу включено показники, які характеризують господарську діяльність бірж. Другий містить структуру пропозицій та угод на біржі. Третій — структуру укладених угод за типом секцій, четвертий розділ відображає котирування цін на біржі. На підставі первинних звітів формується статистичний банк даних за спеціалізацією бірж та в регіональному розрізі.

Статистичний моніторинг економічних явищ і процесів допомагає оперативно управляти не лише цією сферою суспільного життя, а й розробляти державні програми соціально-політичного розвитку.

Соціальний моніторинг: соціологічні і статистичні обстеження

Соціальний моніторинг — це спеціально організована, розгалужена система безперервного й оперативного отримання, опрацювання, зберігання та аналізу даних про актуальні явища і процеси, що відбуваються у суспільстві. На соціальний моніторинг покладається вирішення трьох основних завдань:

- по-перше, інформаційне забезпечення прийняття управлінських рішень в рамках реалізації державної політики (соціально-економічної, демографічної, правової, зовнішньої);
- по-друге, соціальний контроль — перевірка громадянським суспільством порядку виконання владними структурами їхніх безпосередніх зобов'язань, відповідальності, обіцянок;
- по-третє, соціальна діагностика — оперативний аналіз соціального об'єкта з метою визначення його стану, оцінювання проблемної ситуації і перебігу процесу.

Предмет соціального моніторингу становлять суспільні відносини, соціальні явища і процеси та соціальна поведінка членів суспільства.

Об'єкти соціального моніторингу, завдання та головні принципи його організації подано на рисунку 4.2.



Рис. 4.2. Завдання, принципи організації та об'єкти соціального моніторингу

Найважливішим об'єктом соціального моніторингу є населення, а саме соціальні аспекти його природного та механічного руху, соціальна мобільність та соціальна диференціація.

Другий об'єкт — соціальна інфраструктура, зокрема ринок та умови праці, навантаження на вакантні посади, заборгованість із виплати заробітної плати, перебування у вимушених відпусках; сфери: освіти, охорони здоров'я, культури; соціального обслуговування та соціального захисту і допомоги.

З першим і другим об'єктами безпосередньо пов'язані такі:

- соціально-правові відносини з державними інституціями (державотворення та самоврядування, громадянська активність, соціально-політичне напруження тощо);

- умови життя, зокрема рівень і структура споживання товарів та послуг, стан довкілля (атмосферне забруднення, стан водних басейнів та річок, радіаційне забруднення, стан лісів та природно-охоронних і рекреаційних зон); громадська безпека (соціально-правова, політична, військова);

- соціальні настрої і соціальна поведінка (соціально-політичні орієнтації населення, його ставлення до різних суспільно-політичних інституцій, соціальних проблем та окремих подій суспільного життя; електоральна поведінка).

Соціальний моніторинг як узагальнене поняття складається із двох компонентів: соціологічного і статистичного моніторингу.

Соціологічний моніторинг — цілісна система спостережень за суспільними змінами на основі досліджень та аналізу масових уявлень про них.

Соціологічний моніторинг оперує переважно якісною, суб'єктивною інформацією, оскільки має справу із думками респондентів, але на їх підставі визначає загальні об'єктивні тенденції та настрої, що панують у суспільстві.

Основне завдання соціологічного моніторингу — систематичне отримання оперативної (поточної) інформації шляхом опитування населення з найактуальніших проблем і питань, а також щодо подій та явищ, які відбуваються або плануються в суспільстві.

Соціологічний моніторинг виконує роль барометра громадської думки і охоплює такі аспекти: соціально-політичні настрої населення; ставлення громадян до різних суспільно-політичних інституцій; самооцінка рівня життя населення; соціальне самопочуття громадян (їхні думки щодо різних соціальних проблем та окремих подій суспільного життя); електоральна поведінка; рівень соціального напруження та конфліктного потенціалу населення.

Джерелом соціальної інформації поведінково-мотиваційного характеру є матеріали спеціально зорганізованих несудільних (вибіркових, панельних, основного масиву) обстежень, що проводяться громадськими організаціями, державними і недержавними установами, такими як Київський міжнародний інститут соціології (КМІС), Український інститут соціальних досліджень

ім. О. Яременка, Центр О. Разумкова, Центр «Соціальний моніторинг», Центр «Соціальні ініціативи» тощо.

Наприклад, Київський міжнародний інститут соціології проводить дослідження рівня життя населення, стилю життя окремих соціальних груп, якості надання населенню соціальних послуг, вивчає соціальні аспекти впровадження економічних реформ. Акумуляовані в банку даних інституту результати багаторічних обстежень використовуються для моніторингу змін, що відбуваються з об'єктами дослідження.

Моніторинг громадської думки населення України проводиться також консорціумом, до складу якого входить «Український інститут соціальних досліджень імені О. Яременка», Центр «Соціальний моніторинг» та відділ моніторингових досліджень соціально-економічних трансформацій Інституту економіки та прогнозування НАН України. Об'єктами моніторингу громадської думки є: поточні рейтинги політичних сил і окремих політиків; рівень довіри до політичних партій та соціальних інститутів; електоральні й суспільно-політичні настрої; самооцінка матеріального становища та рівня життя, задоволеність життям; питання неформальних і тіньових доходів, корупції; популярність окремих інформаційних джерел тощо. Моніторинг організується як вибіркове обстеження у формі індивідуального інтерв'ю за місцем проживання респондентів. Результати обстеження поширюються на все доросле населення України із прийнятною відносною похибкою репрезентативності.

Для проведення спеціально організованих несудільних обстежень використовують різні способи формування вибіркової сукупності: стратифікований, серійний, квотний добір. Можуть застосовуватися окремі прийоми несудільних обстежень: доступної чи стихійної вибірки, а також прийом «снігового клубка».

Сутність стратифікованого добору полягає в тому, що загальний обсяг вибірки формується пропорційно до структури генеральної сукупності і в кожній страті (типовій групі) здійснюється проста випадкова або механічна вибірка.

За серійної вибірки добираються не окремі респонденти, а однірідні групи або колективи.

Квотний добір застосовується за наявності даних про склад генеральної сукупності за визначальними ознаками. Респонденти добираються цілеспрямовано, із заданими показниками квот. Число показників, які визначають квоти, не має перевищувати чотирьох.

Прийом доступної вибірки застосовують для гіпотетичних генеральних сукупностей, причому обстежуються одиниці за місцем їхнього скупчення (споживачі певних послуг опитуються в пунктах їх надання, члени релігійних конфесій обстежуються в місцях проведення культових заходів).

Прийом «стихійної вибірки» не відповідає принципам випадкового відбору і належить до самоутворювального несучільного добору.

Прийом «снігового клубка» застосовується для невеликої експертної групи, що доповнюється респондентами, обраними за порадою попередніх.

Зазначені способи і прийоми формування вибірових сукупностей докладно описано у підрозділах 5.3, 7.3, 8.1, 9.1, 9.3 пропонуваного навчального посібника.

Усі соціологічні обстеження проводяться шляхом опитування респондентів. Способи опитування можуть бути різними: індивідуальними чи колективними, особистими чи комп'ютеризованими, поштовими чи в онлайн-режимі; вельми популярним є опитування шляхом особистого інтерв'ювання респондентів у формі «face-to-face». Зазвичай це традиційна усна форма, що супроводжується письмовим заповненням опитувальника самим інтерв'юєром. Застосовується й сучасна форма за допомогою електронних пристроїв для відображення форми опитувальника і введення відповідей респондента. Наприклад комп'ютеризоване телефонне опитування (Computer Assisted Telephone Interview — CATI) здійснюються методом квотної вибірки.

Поштові опитування виконуються у формі анкетування респондентів шляхом самореєстрації згідно із наведеною інструкцією. Організуються як обстеження основного масиву, стратифікована або квотна вибірка. Це різновид опитування має певні недоліки, одним з яких є висока частка неповнених анкет та невідповідей в анкетах. Окрім того, часто не дотримується регламент повернення анкет.

Інтернет-опитування здебільшого застосовують для проведення фокус-групових онлайн-дискусій, спостережень і створення спеціалізованих панелей. Організуються як несучільне експертне обстеження чи самоутворювана вибірка. Рідше застосовується квотний добір. Перевагами цього способу опитування є прозорість контролю за процесом обстеження та керованість і спрямування дискусії в заданому руслі. Окрім того, в онлайн-режимі є

можливість повторного звернення до респондента, а також застосування власне випадкового добору.

У практиці соціологічних обстежень використовують так звані «вуличні опитування» — оперативні обстеження респондентів для отримання блиц-інформації з одного-трьох запитань. Частіше організується як самоутворювана вибірка, іноді виконується квотний добір. Цей вид опитування не потребує ретельного розроблення програми спостереження і дає можливість застосувати випадковий добір.

Зазначені способи опитування частіше використовуються для одноразових обстежень. Проте для моніторингу, тобто регулярних спостережень застосовують трекінгові, ролінгові, панельні обстеження та омнібус-опитування.

Трекінгові (моніторингові) обстеження — це регулярні обстеження респондентів, що входять до постійної вибіркової мережі з метою відстежування зміни їхніх оцінок і думок (вибіркові обстеження домогосподарств з питань економічної активності, умов життя та самооцінювання стану здоров'я). Основними перевагами цього виду спостережень є застосування єдиної незмінної програми спостереження та жорстко регламентовані терміни обстеження. Організується обстеження у формі стратифікованої або квотної вибірки чи застосовується метод основного масиву.

Ролінгові обстеження — сучасний різновид моніторингових опитувань у щоденному режимі. Щодня опитується невелика репрезентативна вибірка респондентів, а потім методом «ковзної середньої» поєднуються дані останніх трьох (іноді — п'яти) днів, що дозволяє усунути випадкові відхилення. Важливо, що за цього способу обстеження замовник отримує інформацію про зміну показника, який його цікавить (наприклад, рейтингу партії, політичного діяча) у щоденному режимі і може аналізувати узгодженість між зміною таких рейтингів та характером медійних подій. Ролінгові дослідження проводять за постійною стратифікованою або квотною вибіркою.

Останнім часом дедалі популярнішими стають панельні обстеження. Панельні обстеження — це багаторазові (мінімум дворазове) обстеження одних і тих самих одиниць вибіркової сукупності (респондентів) з метою відстежування зміни їхніх характеристик, умов та поглядів (обстеження умов життя домогосподарств та економічної активності населення). Широко ви-

користовують панельні обстеження в маркетингу для аналізу зміни споживчих намірів. Також цей спосіб застосовують у системі охорони здоров'я для визначення змін у самооцінці населення стану здоров'я, особливо серед категорій, постраждалих унаслідок техногенних катастроф. Надійність результатів панельних обстежень забезпечується єдиною і незмінною програмою спостереження й дотриманням жорстко регламентованих термінів обстеження. Також є можливість підтримання зворотного зв'язку з респондентами. Вибіркове обстеження можна проводити в один із двох способів — за стратифікованою або квотною вибіркою. Іноді використовують методи основного масиву. Статистичні аспекти застосування панельних обстежень розглянуто у параграфі 9.3 нашого навчального посібника.

Омнібус-опитування — регулярне багатоцільове опитування населення, проведене на замовлення одночасно кількох замовників з різними завданнями.

Кожен замовник включає в анкету й платить тільки за ті запитання, які безпосередньо пов'язані з його інформаційними потребами, а основні соціально-демографічні характеристики використовуються спільно всіма учасниками проекту на безоплатній основі. Головними перевагами цього опитування є швидкий і зручний спосіб одержання якісної й надійної інформації, а також економія витрат замовника. Організується омнібус-опитування як несудільне експертне обстеження або за допомогою квотної вибірки.

Для більшості згаданих способів обстеження використовується такий статистичний інструментарій, як анкета чи опитувальний листок (для інтерв'ю). До них зазвичай додається інструкція щодо заповнення у вигляді окремого документа або як структурний елемент анкети (опитувальника), розміщується на початку чи наприкінці.

Зазвичай анкета складається з двох частин — адресної і власне переліку запитань. Адресна частина містить звернення до респондента із зазначенням мети опитування і проханням висловити власну думку. Запитання анкети можуть бути закритого, напівзакритого та відкритого типу.

Запитання *закритого типу* — до яких додається повний набір відповідей, а респондентові пропонується лише обрати потрібні. Іноді кількість можливих відповідей обмежується (не більше або не менше двох-трьох відповідей).

Запитання *напівзакритого типу* містять перелік готових відповідей, а також вільний рядок для самовизначення.

Запитання *відкритого типу* потребують самостійного формулювання відповіді.

При організації масових анкетних спостережень застосовують переважно запитання закритого та напівзакритого типу. Вони дають змогу, по-перше, отримати адекватні відповіді, а по-друге, формалізувати відповіді, накопичити їх та кількісно обробити. Практика показує, що оптимальна кількість запитань в анкеті варіює від 12-ти до 15-ти. При розробленні лаконічних і зрозумілих відповідей намагаються дотримуватися певного порядку їх розташування — найбільш імовірні містяться наприкінці переліку відповідей.

Кількісний опис відповідей в анкеті здійснюється за певною шкалою. Для опрацювання даних анкет застосовують три типи шкал: номінальний, порядковий, метричний.

Номінальна шкала характеризує належність одиниць до відповідної групи властивості, причому послідовність розміщення груп значення не має. Кожне значення (відповідь) оцифровується як альтернативна ознака. Альтернативна ознака набуває двох протилежних значень (1 — є відповідь, 0 — немає).

В анкетах застосовують два різновиди номінальної шкали: альтернативна; набір ознак (перелік незалежних відповідей).

Порядкова (рангова) шкала характеризує послідовність прояву властивості, окремі градації описуються рангами, балами (відмінно, добре, задовільно, погано / подобається, не дуже подобається, не подобається).

Метрична шкала визначає (описує) розмір, кількість та рівень властивості на підставі безпосереднього вимірювання та підрахунку.

Найбільш інформативною й універсальною з точки зору статистичного опрацювання даних опитування є метрична шкала. За ознаками метричної шкали можна не лише побудувати різні статистичні групування (структурні, типологічні, аналітичні, комбінаційні, комбінаційно-аналітичні), а й обчислити узагальнюючі характеристики (середній, модальний, медіанний рівень), розрахувати показники варіації, диференціації, форми розподілу (асиметрії, ексцесу), а також перевірити статистичні гіпотези. Для визначальних характеристик можна обчислити похибки репрезентативності й поширити результати обстеження на генеральну сукупність респондентів.

Нижчою за рівнем інформативності є порядкова шкала. Для ознак порядкової шкали можна:

- побудувати структурні, типологічні та комбінаційні групування;
- виконати частотний аналіз (частка відповідей на запитання у процентах до чисельності респондентів, а також до загальної кількості тих, хто надав відповідь) і визначити порядкові середні (моду, медіану та інші квантили);
- обчислити дисперсії для частки окремих відповідей;
- розрахувати показники співзалежності відповідей;
- обчислити середній центрований ранг.

Середній центрований ранг розраховується за середньою арифметичною зваженою центрованих рангів R^0_j ; кожної j -ої групи відповідей на частку d_j — ої групи відповідей респондентів:

$$\underline{R} = \frac{\sum_{j=1}^m R^0_j \cdot d_j}{\sum_{j=1}^m d_j} \quad (4.1)$$

Центровані ранги R^0_j встановлюються в такому порядку: центральному рангу надається цифра «0», а решті — додатні та від'ємні порядкові номери, додатні — позитивним градаціям, від'ємні — негативним. Так, для непарної шкали: +2; +1; 0; -1; -2, а для парної: +2; +1; -1; -2.

На підставі порядкової шкали за порядковими (нецентрованими) значеннями можна обчислити нормований коефіцієнт якості K_N , що характеризує інтенсивність прояву властивості і набуває значень від 0 до 100% й обчислюється за формулою:

$$K_N = \frac{\bar{x} - \bar{x}_0}{0,5 \cdot R} \cdot 100, \quad (4.2)$$

де $\bar{x} = \sum_{j=1}^m x_j \cdot d_j$ — середньозважений порядковий ранг,

\bar{x}_0 — середній розмір шкали; розмах шкали $R = x_{\max} - x_{\min}$.

Порядок обчислення центрованого рангу і нормованого коефіцієнта якості розглянемо на прикладі результатів опитування громадян України у 2015 році щодо міри довіри до Збройних сил України (табл. 4.1).

Таблиця 4.1

**РОЗПОДІЛ ВІДПОВІДЕЙ ГРОМАДЯН УКРАЇНИ
ЩОДО ДОВІРИ ДО ЗБРОЙНИХ СИЛ УКРАЇНИ**

Міра довіри	Частка респондентів, % d_j	Порядковий ранг, x_j	Центрований ранг, R_j^0
Цілком довіряю	61	3	+1
Частково довіряю	18	2	0
Зовсім не довіряю	21	1	-1
Разом	100	x	x

Центрований середній ранг становить $\bar{R}^0 = 1 \cdot 0,61 + 0 \cdot 0,18 + (-1) \cdot 0,21 = 0,40$ бала, а нормований коефіцієнт якості $K_N = \frac{2,4 - 2}{0,5 \cdot 2} \cdot 100 = 40\%$, де $\bar{x} = \sum_1^m x_j \cdot d_j = 3 \cdot 0,61 + 2 \cdot 0,18 + 1 \cdot 0,21 = 2,4$ бала, $\bar{x}_0 = (3 + 2 + 1) : 3 = 2$, $R = 3 - 1 = 2$. Тобто міра довіри громадян до Збройних сил України дорівнює 40 %.

Найменш інформативними є ознаки номінальної шкали, для яких можна:

- побудувати структурні, типологічні та комбінаційні групування;
- виконати частотний аналіз (частка відповідей на запитання у процентах до чисельності респондентів, а також до загальної кількості тих, хто надав відповідь);
- обчислити середню кількість зазначених відповідей на одне запитання (сумарна частка відповідей на запитання, що дозволяють обрати кілька відповідей, може бути більшою за 100 %, тобто на одне запитання респонденти надали більше однієї відповіді);
- розрахувати дисперсію для частки окремих відповідей.

Статистичний моніторинг — це спеціально організоване комплексне неперервне відстежування стану суспільного середовища та соціально-політичних процесів у ньому із періодичною реєстрацією сталого набору його кількісних та якісних параметрів чи індикаторів. На відміну від соціологічного статистичний моніторинг передбачає, у першу чергу, реєстрацію кількісних показників соціально-політичного середовища станом на окремі дати чи за інтервали часу.

На підставі зібраної й систематизованої інформації аналізується динаміка показників та здійснюється прогнозування їхніх рівнів. У рамках статистичного моніторингу може виконуватися імітаційне моделювання показників об'єкта спостереження та оцінювання можливих «ризиків» або очікуваних позитивних (реальних) змін. Результати статистичного моніторингу застосовують як підґрунтя для оперативного або стратегічного управління об'єктом спостереження. Також вони використовуються як інформаційне забезпечення інших видів спостереження (основа для складання статистичних реєстрів, обліків).

Своєю чергою, статистичний моніторинг є комбінацією різних форм, видів і способів спостереження. Так, в якості фундаментальної і надійної інформаційної основи моніторингу використовують матеріали перепису населення (чисельність та структура населення за демографічними, соціальними та економічними характеристиками). Для відстежування змін у міжпереписний період використовують дані суцільного поточного обліку (реєстрація демографічних подій).

Окремим компонентом моніторингу є статистичні реєстри (Державний демографічний реєстр, реєстр виборців тощо). Інформацію щодо різних аспектів функціонування соціальної інфраструктури надає статистична звітність міністерств, відомств та установ відповідної сфери (охорони здоров'я, фізичної культури та спорту, освіти, навколишнього середовища, служб соціального захисту та допомоги тощо).

Неодмінним інформаційним ресурсом є дані регулярних вибірових обстежень домогосподарств, на постійній основі (стосовно умов життя й економічної активності) або спеціально організованих тематичних несучільних обстежень, у тому числі соціологічних опитувань. Розглянемо кілька прикладів застосування статистичного моніторингу у соціальній сфері.

Статистичний моніторинг у системі охорони здоров'я ведеться з метою поліпшення надання пацієнтам медичної допомоги, виявлення проблем та порушень і проведення організаційних заходів (корекції умов, ресурсного забезпечення, дій медичного персоналу), спрямованих на поліпшення якості медичної допомоги. Моніторинг здійснюється щорічно всіма первинними клінічними закладами охорони здоров'я.

Результати моніторингу узагальнюються державним підприємством «Державний експертний центр Міністерства охорони

здоров'я України» і подаються у формі річного звіту до Міністерства охорони здоров'я України, структурних підрозділів з питань охорони здоров'я не пізніше 01 червня року, наступного за звітним періодом.

Оцінювання результатів моніторингу здійснює державне підприємств «Державний експертний центр Міністерства охорони здоров'я України» шляхом повного або вибіркового опрацювання джерел медичної інформації, включених до відповідних індикаторів якості медичної допомоги (із або без використання засобів автоматизованого опрацювання інформації).

Статистичним інструментарієм такого моніторингу є «Форма аудиту результатів моніторингу клінічних індикаторів якості медичної допомоги пацієнту», що містить ідентифікаційні ознаки пацієнта (стать, рік народження, соціальний стан). Окремо вказуються медичні характеристики (назва локального протоколу медичної допомоги, діагноз). У тому числі зазначаються причини невиконання протоколу медичної допомоги: пізні звернення пацієнта, неможливість транспортування пацієнта, відсутність консультанта необхідної спеціалізації. Також реєструються причини відмови пацієнта від допомоги: індивідуальна непереносимість лікарського засобу, відсутність своєчасного доступу до послуги або її відсутність у закладі охорони здоров'я.

У рамках реалізації медичної реформи в Україні створено *Медичну інформаційну систему (МІС)* — систему автоматизованого документообігу для медичних закладів, у якій об'єднані порядок прийняття медичних рішень, електронні медичні картки, моніторинг стану пацієнта (з діагностичного обладнання), фінансову та адміністративну інформацію.

За новими правилами будь-який медичний заклад відтепер має взаємодіяти з центральним компонентом *e-Health* через МІС, котрий виконує роль управлінського центру лікарні. *E-Health* — електронна медична система реєстрації опрацювання та зберігання даних про пацієнтів та лікарів первинної ланки, яку запроваджено у 2017 році. Ця система, з одного боку покликана спростити процедуру реєстрації пацієнтів і запису їх до сімейних лікарів у режимі online або за телефоном. З іншого боку — має контролювати ефективність використання коштів, що виділяються на охорону здоров'я (370 гривень на рік за кожного пацієнта). У системі *e-Health* ведеться поточний облік кількості пацієнтів, які закріплені за конкретним лікарем. Від-

повідно робота лікаря оплачуватиметься залежно від його завантаженості пацієнтами.

Моніторинг стану здоров'я населення здійснюється в процесі вибіркового обстежень домогосподарств з питань умов життя. Реєструються такі характеристики стану здоров'я населення:

1. Самооцінка загального стану здоров'я.
2. Відчуття відсутності/наявності фізичних хвороб.
3. Відчуття психоемоційного дискомфорту/благополуччя.
4. Відсутність/наявність обмежень працездатності та життєвої активності.

5. Відповідність медико-біологічним нормам індивідуальних показників здоров'я (артеріальний тиск, формула крові, маса тіла).

Результати цих обстежень відстежуються в часі і порівнюються у територіальному розрізі, що дозволяє не лише мати загальну та регіональну картину стану здоров'я населення, а й вживати цілеспрямованих управлінських заходів у плані його поліпшення й надання ефективної допомоги.

Статистичний моніторинг у сфері освіти здійснюється за для оцінювання розвитку системи, ефективності впровадження освітніх реформ та визначення якості підготовки випускників закладів освіти.

Моніторинг ґрунтується на матеріалах перепису населення щодо рівня грамотності та рівня освіти населення з урахуванням соціально-демографічних характеристик (стать, вік, громадянство, вільне володіння мовою, вид поселення, статус зайнятості, вид діяльності, джерела засобів існування).

Залучаються також дані щорічної статистичної звітності: «Звіт дошкільного навчального закладу» (№ 85-к); «Звіт про кількість дітей шкільного віку» (6–18 років) (№ 77рвк); «Звіт ЗВО на початок навчального року» (№ 2–3 нк); «Звіт про роботу аспірантури та докторантури» (№ 1-нк).

Також використовуються адміністративні дані міністерства освіти і науки (МОН) України. Це зведені звітні дані про діяльність дошкільних закладів, дитячих будинків, ПТНЗ, про кількість педагогічних працівників, кількість працівників, які проходили профнавчання і перепідготовку, про забезпеченість загальноосвітніх навчальних закладів комп'ютерною технікою та доступом до інтернету. Утім, більшість показників, які реєструються звітністю, є лише формальними кількісними характеристиками стану та розвитку системи освіти в країні.

Для оцінювання ефективності вищої освіти здійснюєть *моніторинг якості освіти* за допомогою рейтингів закладів вищої освіти. Використовують одночасно кілька рейтингів: офіційний МОН України; комерційні, міжнародні та консолідований рейтинг закладів вищої освіти України.

Офіційний рейтинг МОН України відбиває лише рівень матеріально-технічного та кадрового забезпечення закладів вищої освіти (ЗВО). Серед комерційних рейтингів — проект «Компас», рейтинг «Вебометрик» та рейтинг журналу «Деньги». Міжнародні рейтинги: QS World University Ranking відстежує лише 400 найкращих університетів світу; рейтинг ЮНЕСКО «ТОП-200 Україна».

Проект «Компас» — загальноукраїнський рейтинг, що враховує думку роботодавців, про відповідність знань і навичок випускників вітчизняних ЗВО потребами ринку праці. Одиницями спостереження є представники компаній-роботодавців, експерти, випускники ЗВО, які є об'єктом моніторингу.

Обстеження представників компаній-роботодавців проводиться у першому кварталі календарного року і зорганізується у формі стратифікованої вибірки за такими показниками: розмір підприємства за кількістю працівників, вид економічної діяльності, належність до територіально-адміністративних одиниць України. Основу вибірки становлять великі і середні компанії-роботодавців, які належать до одного із 10 основних видів економічної діяльності. Сукупність експертів формують керівники відділів кадрів та HR-відділів (менеджери з персоналу і рекрутери), а також керівники провідних компаній за одним із видів економічної діяльності. Спосіб обстеження — особисте, телефонне, поштове (електронна пошта) та онлайн-опитування. Аналогічно опитуються випускники ЗВО денної форми навчання за одним із 5-ти пріоритетних напрямів підготовки за освітнім рівнем бакалавр і магістр.

Рейтинг журналу «Деньги» відстежує якість підготовки випускників лише з точки зору вартісного оцінювання їхньої праці (заробітної плати).

Рейтинг ЮНЕСКО «ТОП-200 Україна» складається для 200 кращих ЗВО країни на підставі статистичних даних та експертних оцінок. Позицію закладу вищої освіти визначає інтегрований індекс, складовими якого є три індекси: якості науково-педагогічного потенціалу, якості навчання, індекс міжнародного визнання.

Консолідований рейтинг вищів України включає 275 ЗВО III–IV рівнів акредитації, а також 4 додаткові рейтинги у номінаціях:

«Кращі класичні університети України», «Кращі приватні ЗВО України», «Кращі київські ЗВО» та «Кращі ЗВО регіонів України».

Консолідований рейтинг складається шляхом узагальнення авторитетних рейтингів: ЮНЕСКО “Топ-200 Україна”, “Компас”, “Вебометрикс”, кожен з яких використовує різні критерії оцінювання навчальних закладів вищої освіти.

Таким чином, статистичний моніторинг, що виконується в рамках соціального моніторингу, з одного боку, дає важливу й оперативну інформацію для управління соціальними процесами на загальнодержавному та регіональному рівнях, а з іншого — слугує індикатором ефективності прийняття управлінських рішень, реалізації реформ і втілення державної політики розвитку країни.

Вправи та питання для самоконтролю

1. Зазначте особливості моніторингу як виду статистичного спостереження.

2. Назвіть сфери використання сучасних моніторингових систем.

3. З’ясуйте аналітичні можливості даних сайта Держстату України для здійснення моніторингу підприємницької діяльності.

4. У який спосіб дані офіційної статистики дозволяють провести моніторинг запровадження реформ у системі освіти і охорони здоров’я?

5. Визначте вид моніторингу за метою організації, специфікою даних і рівнем узагальнення для таких обстежень:

1) якості капітального ремонту дорожньо-транспортного покриття доріг загальнодержавного значення;

2) споживчих намірів населення та конкурентного середовища при запровадженні електронної системи паркування легкових автомобілів у місті;

3) рівня цін на первинну житлову нерухомість різного класу від усіх забудовників регіону.

6. Визначте вид моніторингу за специфікою даних, часом проведення і рівнем узагальнення для таких обстежень:

1) поштових і банківських переказів з метою контролю руху коштів, отриманих злочинним шляхом;

2) контролю якості питної води в артезіанських свердловинах міста;

3) думки населення з 35-ти країн світу, які є користувачами соціальної мережі Twitter у рамках молодіжного проекту U-Report щодо соціального благополуччя молоді.

7. Складіть план статистичного моніторингу закупівель товарів та послуг за державні кошти через відкриті конкурсні торги та за допомогою тендерних процедур.

8. Складіть план статистичного моніторингу щеплення та вакцинації дошкільнят.

9. Опишіть форми, види і способи статистичного спостереження, які можуть використовуватися для моніторингу заборгованості населення зі сплати комунальних послуг.

10. Розробіть програмно-методологічні й організаційні питання плану моніторингу подання електронних петицій.

11. Розробіть програмно-методологічні й організаційні питання плану моніторингу процесів децентралізації та місцевого самоврядування.

12. Розробіть статистичний інструментарій для проведення соціологічного обстеження студентів бакалаврського освітнього рівня щодо їхніх подальших освітніх намірів.

13. Розробіть статистичний інструментарій для проведення моніторингу утворення та утилізації побутових відходів.

14. Опишіть структуру статистичного банку даних для моніторингу осіб, звільнених з посади держслужбовця у зв'язку із виконанням процедури люстрації.

15. Опишіть структуру статистичного банку даних для моніторингу учасників фондового ринку та емісії цінних паперів.

16. У результаті онлайн-опитування студентів щодо їхнього ставлення до соціальних мереж одержано відповіді на такі запитання:

Чи вважаєте Ви себе залежним від соціальних мереж?	Відповіді респондентів, у % до підсумку
Ні, не вважаю	31
Залежу частково	56
Дуже залежу	13
Загалом	100

Якби соціальні мережі мали платний доступ, Ви б ними користувалися?	Відповіді респондентів, у % до підсумку
Так, за будь-яких умов	7
Так, регулярно за помірну ціну	36
Зрідка, за необхідності	45
Не користуватимуся	12
Загалом	100

Для відповідей на кожне запитання обчисліть середній центрований ранг і нормований коефіцієнт якості. Зробіть висновки щодо міри залежності опитаних студентів від соціальних мереж.

Рекомендована література

1. Балабанов Г.В., Вишневський В.В. Територіальний соціально-економічний моніторинг в Україні / Ін-т географії НАН України. — К. : Нора-Друк, 2001. — 58 с.
2. Галіцин В.К. Системи моніторингу. — К. : КНЕУ, 2000. — 232 с.
3. Концептуальні основи статистичного моніторингу : Монографія / За ред. М.В. Пугачової. — К.: ІВЦ Держкомстату України, 2003. — 343 с.
4. Осауленко О.Г. Проект «Розвиток системи державної статистики для моніторингу соціально-економічних перетворень в Україні» та його значення в подальшій розбудові національної статистики // Статистика України. — 2004. — № 3. — С. 4–6.
5. Пальян З.О. Демографія : Навч. Посібник. — К. : Видавничо-поліграфічний центр «Київський університет», 2014. — 222 с.
6. Пугачова М.В. Методологічні засади статистичного моніторингу ділової активності підприємств : Монографія. — К. : ДП Інформ.-анал. агентство, 2007. — 535 с.
7. Системи фінансового моніторингу. Методологія проектування / За ред. О.В. Мозенкова. — Харків: ВД «ІНЖЕК», 2005. — 151 с.

Розділ 5

ВИБІРКОВІ ОБСТЕЖЕННЯ: ЗАВДАННЯ І СПОСОБИ ОРГАНІЗАЦІЇ

Сутність, переваги і сфери використання вибіркового обстежень

Основа вибірки, типи основ

Способи формування вибіркового сукупностей

Застосування невивпадкового (спрямованого) добору

Після вивчення цього розділу студент має

- *знати*: сутність і сферу використання вибіркового обстеження; способи формування вибіркового сукупностей; переваги і недоліки кожного способу; вимоги до формування вибірки щодо надійності результатів;
- *уміти*: формувати основу вибірки з урахуванням якості наявних інформаційних джерел; формулювати вимоги до результатів вибіркового обстеження; знаходити компроміс між витратами і точністю результатів обстеження; обрати найбільш ефективний спосіб організації вибірки.

Сутність, переваги і сфери використання вибіркового обстеження

На етапі переходу до інформаційного суспільства зростає попит на статистичну інформацію, змінюється характер вимог до неї — від фіксації певних фактів до всебічного аналізу закономірностей функціонування і розвитку об'єктів управління. Задовольнити зростання попиту на інформацію екстенсивним шляхом, тобто за рахунок збільшення обсягів статистичних даних, нереально, а отже, необхідна інтенсифікація процесів збирання й опрацювання даних, застосування ефективніших технологій. У вирішенні названої проблеми важливе місце посідають вибіркового обстеження. Невпинно розширюються сфера використання вибіркового обстеження, удосконалюються методи і засоби забезпечення надійності їхніх результатів.

Сутність вибірки як різновиду статистичного спостереження полягає в тому, що обстеженню підлягає не вся первинна сукуп-

ність, а лише певним способом дібрана частина сукупності. За результатами обстеження дібраної частини судять про структуру й основні параметри первинної сукупності.

Практика вибіркових обстежень досить різноманітна. Це обстеження домогосподарств, маркетингові дослідження, аудиторські перевірки великих фірм, вивчення громадської думки тощо. Підстав для використання вибіркових обстежень багато. Безперечною перевагою вибіркових обстежень, є економія матеріальних, трудових і фінансових ресурсів. У разі обстеження невеликої частини сукупності зменшуються помилки реєстрації, можна розширити й деталізувати програму обстеження. В умовах швидкоплинної соціально-економічної ситуації не менш важливе значення має фактор часу — скорочуються терміни отримання необхідної статистичної інформації.

При вивченні певного кола соціально-економічних явищ вибіркве спостереження єдино можливе. Це стосується передусім перевірки якості продукції, якщо вона пов'язана із руйнуванням зразків (термін служби електроламп, калорійність палива тощо). Часом вибіркве спостереження поєднується із суцільним. Наприклад, під час перепису населення кожна четверта одиниця спостереження дає докладнішу інформацію. Крім того, вибірквий метод використовують для прискореного опрацювання матеріалів суцільного спостереження та перевірки правильності даних переписів і одноразових обстежень. Дібрати дані не про всі, а лише про частину сукупності простіше, швидше, дешевше, але дібрана частина репрезентуватиме ціле з прийнятним ступенем точності лише за певних умов.

У світі накопичено неабиякий досвід організації вибіркових обстежень, розроблено системи методологічних принципів і способів формування вибіркових сукупностей, правил і методів оцінювання параметрів об'єктів дослідження за неповними даними, поширення результатів вибіркового обстеження на генеральну сукупність. Усі вони спрямовані на забезпечення репрезентативності результатів обстежень.

Невпинно розширюється практика вибіркових обстежень в Україні. В державній статистиці традиційно найбільш масштабними є базові вибіркові обстеження населення (домогосподарств). Це одне з основних джерел інформації, необхідної для всебічного аналізу рівня життя населення, моніторингу бідності, комплексного оцінювання стану ринку праці, для вимірювання

інших соціально-економічних явищ і процесів у контексті макро-економічних розрахунків.

Органи державної статистики здійснюють також моніторинг споживчих цін і цін виробників, проводять обстеження трудової міграції, медико-демографічні обстеження, інші обстеження та одноразові обліки відповідно до міжнародних стандартів і потреб користувачів інформації усіх рівнів.

Створення ЄДРПО, в якому фіксуються всі суб'єкти господарювання незалежно від форми власності, вможливило проведення різноманітних вибіркового обстежень в економіці. Велике значення мають вибіркові обстеження на рівні окремих підприємств:

- у промисловості для статистичного контролю якості продукції, для оцінювання використання устаткування, робочого часу, ефективності нових технологій;
- в аграрному секторі при оцінюванні стану посівів, урожайності, продуктивності худоби, втрат;
- у торгівлі маркетингові дослідження попиту і споживчих реакцій на окремі товари, ефективності реклами тощо.

Досить швидких обертів набирають вибіркові обстеження ділової активності підприємств (кон'юнктурні обстеження), монетарні обстеження в аудиті, медіаметричні вимірювання теле- та інтернет-аудиторії, моніторинг перебігу актуальних для суспільства процесів. Ширшає сфера використання соціологічних опитувань при оцінюванні соціально-політичної ситуації і соціального напруження, виявлення електоральних настроїв тощо.

В організації вибіркового обстеження вирізняють генеральну і вибірку сукупності. *Генеральною* називають сукупність одиниць об'єкта дослідження, яким властиві цільові характеристики, що підлягають реєстрації; *вибірковою* — дібрану для обстеження частину генеральної сукупності. Генеральна сукупність може бути *реальною* або *гіпотетичною*. *Реальна* генеральна сукупність — це множина об'єктів, подій, зразків або сукупність вимірів, яка відповідно до мети дослідження має свої просторові й часові межі; *гіпотетична* сукупність оперує не кількістю реальних елементів, а кількістю всіх можливих варіантів випадкових величин. Дібрана за певними правилами вибірка сукупність завжди репрезентує генеральну, а тому всі її статистичні характеристики (середня, дисперсія, частка одиниць сукупності з певним значенням ознаки) розглядаються як оцінювання відповідних характеристик генеральної сукупності (рис. 5.1).

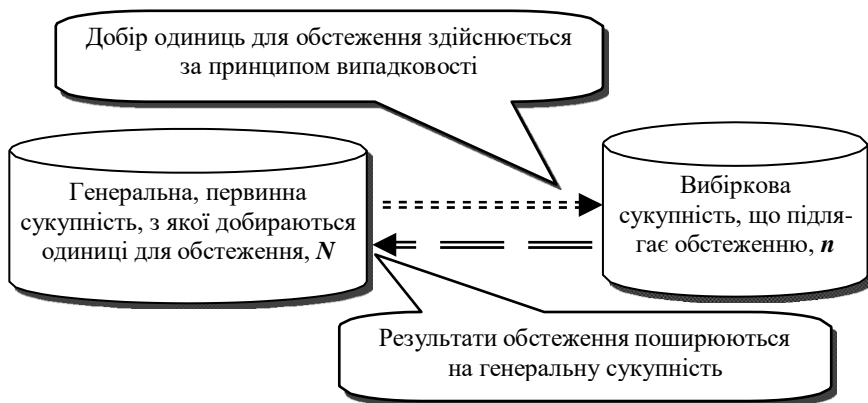


Рис. 5.1. Генеральна і вибірка сукупності

Оскільки вибірка сукупність не точно відтворює склад генеральної сукупності, то й вибіркові оцінювання не збігаються з відповідними характеристиками генеральної сукупності. Розбіжності між ними називають *похибками репрезентативності* A : для середньої — це різниця між генеральною \bar{x}_0 та вибірковою \bar{x} середніми, для частки — різниця між генеральною p_0 і вибірковою p частками тощо.

За причинами виникнення похибки репрезентативності поділяються на систематичні та випадкові: $A = \lambda + \Delta$. *Систематичні* (тенденційні) похибки λ виникають, коли при формуванні вибіркової сукупності порушено принцип випадковості добору (упереджений добір одиниць, недосконала основа вибірки, нерівноцінна заміна одиниць у процесі спостереження тощо). Ці похибки для всіх одиниць сукупності односпрямовані і призводять до зміщення результатів обстеження. *Випадкові* похибки Δ — це наслідок випадковості добору елементів для дослідження і пов'язаних з цим розбіжностей між структурами вибіркової та генеральної сукупностей.

При організації вибіркового обстеження важливо уникнути систематичних похибок. Це можливо за умови неупередженого, об'єктивного добору одиниць для обстеження, незалежності добору від значень цільових ознак. Щодо притаманних вибіркового спостереженню випадкових похибок, то їх уникнути неможливо,

але вони не мають тенденційного характеру й не ведуть до зміщення оцінок; теорія вибіркового методу дає математичну основу для обчислення випадкових похибок з достатньою точністю та надійністю.

Об'єктивною гарантією того, що вибірка репрезентує всю генеральну сукупність, є дотримання наукових принципів організації та проведення спостережень. Визначальний принцип організації вибіркового спостереження — випадковість добору, який забезпечує усім одиницям генеральної сукупності рівні шанси потрапити у вибірку; таку вибірку визначають як рандомізовану. Якщо генеральна сукупність містить N елементів, а для обстеження потрібно вибрати з них частину n , то кількість можливих вибірок становить:

$$C_N^n = \frac{N!}{n!(N-n)!}.$$

Усі вони мають однакову ймовірність $\frac{1}{C_N^n}$ потрапити до вибірки, але кожна з них містить певну похибку, що відбиває факт випадковості добору.

Згідно з граничними теоремами закону великих чисел, який є теоретичною підґрунтям вибіркового методу, в масових явищах вплив різних випадкових причин на формування закономірностей та узагальнених характеристик сукупності надзвичайно малий або практично не залежить від випадку. Оскільки похибка вибірки виникає як наслідок випадкових розбіжностей між вибірковою і генеральною сукупностями, то при достатньо великому обсязі вибірки вибіркові характеристики наближаються (сходяться за ймовірністю) до відповідних параметрів генеральної сукупності. Отже, спираючись на доведення граничних теорем, можна стверджувати, що вибіркові сукупності здатні відтворювати, репрезентувати основні характеристики генеральної сукупності.

За умови необмеженої кількості вибірок однакового обсягу з однієї й тієї самої генеральної сукупності розподіл вибірових середніх асимптотично наближається до нормального, більшість значень вибірових середніх, незалежно від розподілу генеральної сукупності, зосереджується навколо генеральної середньої, а

отже, найбільшу ймовірність мають відхилення, близькі до нуля. Чим більше відхилення, тим менша його ймовірність. Для будь-якої ймовірності існує межа відхилень вибіркової середньої від генеральної. Застосовуючи математичний апарат теорії ймовірностей, для однієї конкретної вибірки можна визначити:

- похибку вибірки з узятою ймовірністю;
- ймовірність того, що похибка вибірки для середньої (частки) не перевищить допустиму величину Δ , тобто $|\bar{x} - x_0| \leq \Delta$ чи $|p - p_0| \leq \Delta$;
- мінімально необхідний обсяг вибірки, який забезпечить потрібну точність результатів із узятою ймовірністю.

Отже, при вибіркового спостереженні порівняно із суцільним зменшуються обсяг спостережень і помилки реєстрації, проте з'являється похибка репрезентативності. Об'єктивною гарантією того, що вибіркові характеристики будуть репрезентативними, тобто повно й адекватно відбиватимуть властивості та пропорції генеральної сукупності, є дотримання наукових принципів організації та проведення спостережень, неупередженого підходу до добору одиниць для обстеження, зокрема:

- випадковість добору, яка забезпечує всім одиницям генеральної сукупності рівні можливості потрапити до вибірки;
- пропорційне представництво окремих сегментів сукупності за умови значного її розшарування.

Саме система правил добору одиниць сукупності для обстеження, способів оцінювання точності результатів обстеження та поширення їх на генеральну сукупність становить сутність вибіркового методу.

Будь-який проект з організації вибіркових обстежень чи моніторингу соціально-економічних процесів унікальний, має часові рамки і відповідно до масштабу, змісту і термінів виконання потребує певних ресурсів. У процесі проектуванні вибіркового обстеження важливо забезпечити рівновагу між статистичною ефективністю (точністю результатів) та ефективністю витрат на обстеження, тому увага концентрується на таких аспектах, як: формування ефективної основи вибірки; обґрунтування способу добору одиниць сукупності для обстеження (дизайну вибірки); визначення мінімально достатнього обсягу вибіркової сукупності й обґрунтування найоптимальнішого способу розміщення вибірки з урахуванням витрат на обстеження.

Вибіркове обстеження здійснюється шляхом планування та реалізації взаємопов'язаних програмно-методологічних, організаційних і технологічних процедур, а саме:

- визначення мети обстеження, об'єкта та одиниці спостереження;
- розроблення програми і дизайну вибірки;
- визначення необхідного обсягу матеріальних та фінансових ресурсів;
- формування вибірки і проведення польового етапу обстеження;
- опрацювання результатів обстеження і поширення їх на генеральну сукупність.

Відповідно до цілей вибіркового обстеження на етапі його проектування визначають вимоги стосовно розроблення технічного завдання та ефективного дизайну вибірки. Спільними для всіх обстежень є вимоги щодо:

- а) надійності результатів вибірових обстежень;
- б) урахування очікуваного коефіцієнта відповідей.

Одночасно визначають основу вибірки, окреслюють перелік характеристик, яким мають задовольняти одиниці сукупності, здійснюють аналіз наявної інформаційної бази. Послідовність процедур формування вибіркової сукупності ілюструє рис. 5.2.

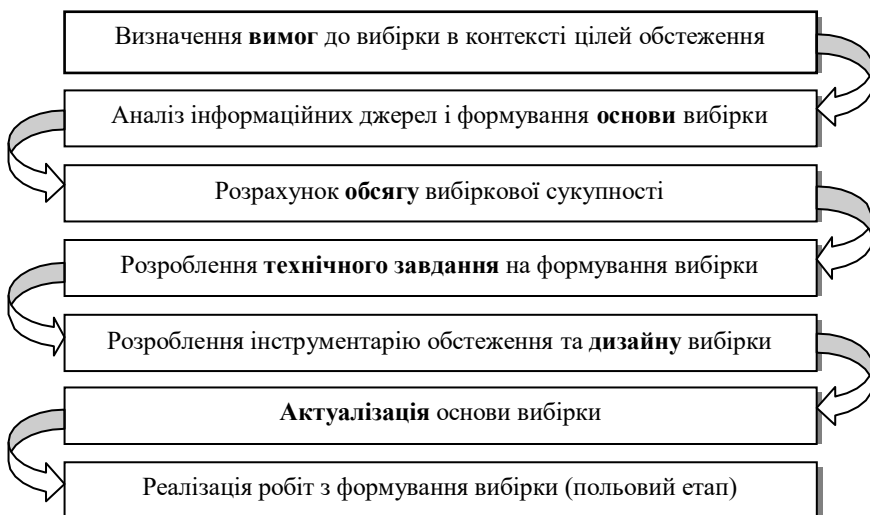


Рис. 5.2. Послідовність процедур формування вибіркової сукупності

На етапі формування основи вибірки визначають об'єкт дослідження, чітко встановлюють характеристики, яким мають задовольняти одиниці сукупності і джерела інформації. Важливим етапом підготовки та проведення будь-якого вибіркового обстеження є обґрунтування ефективної організаційно-логічної моделі структури вибіркової сукупності та принципів її формування — дизайну вибірки. Аби забезпечити рівновагу між точністю результатів і ефективністю витрат на обстеження, визначають мінімально достатній обсяг вибіркової сукупності й обґрунтовують економічно ефективний спосіб розміщення вибірки з урахуванням витрат на обстеження. Окремі етапи вибіркового обстеження за необхідності можуть бути апробовані в пілотному обстеженні. Передусім це стосується програми обстеження.

Основні параметри вибіркової сукупності та узгоджені вимоги щодо організації обстеження і поширення його результатів на генеральну сукупність визначаються технічним завданням на формування вибірки. У технічному завданні формулюється мета вибіркового обстеження, вказуються генеральна сукупність та одиниця спостереження, обсяг і термін дії вибірки, наявність та схема ротації одиниць спостереження; тип вибірки, способи та одиниці добору; інформаційні джерела для формування вибірки; вимоги щодо надійності оцінок основних показників; очікуваний коефіцієнт відповідей респондентів, а також початкове навантаження на одного статистика з проведення обстеження.

Зібрані первинні дані опрацьовують, аналізують, оцінюють надійність результатів і поширюють на генеральну сукупність. Європейська практика гарантування якості статистичних продуктів, у тому числі результатів вибірових обстежень, передбачає складання стандартних звітів з якості. Такі звіти мають уніфіковану структуру, чітко визначений набір критеріїв якості та їхніх індикаторів і характеристик. У Додатку В 3 наведено індикатори якості статистичних продуктів, рекомендовані Євростатом для відображення у звітах з якості [Методологічні положення, 2018].

Основа вибірки, типи основ

Основою вибірки (*sampling frame*) називають упорядковану за певними ознаками множину одиниць первинної сукупності, з якої за встановленою схемою здійснюють добір одиниць для обстеження. При проектуванні вибіркового обстеження обґрунтування основи вибірки має вирішальне значення, адже саме основа визначає об'єкт дослідження і його структуру, що впливає на ефективність обстеження і надійність результатів.

При формуванні основи вибірки вирізняють поняття одиниці добору і одиниці вибірки (спостереження). *Одиниця добору* — це одиниця генеральної сукупності, яка при реалізації процедури добору з певною ймовірністю може бути включена до вибірки. Тобто основу вибірки можна визначити як перелік, список одиниць добору. *Одиниця вибірки* — це елемент сформованої вибіркової сукупності, який є носієм ознак, що підлягають реєстрації в процесі спостереження (цільових ознак). Часто одиниця добору і одиниця вибірки збігаються (обстеження малих підприємств, перевірка якості штучних товарів тощо). У такому разі основа вибірки адекватна генеральній сукупності. Коли добір здійснюють серіями (гніздами) одиниць, а спостереженню піддаються окремі одиниці серії, тоді одиниця спостереження і одиниця добору не збігаються. Наприклад, в обстеженні економічної активності населення одиницями добору є домогосподарства, а одиницями спостереження — особи віком 15–70 років, котрі є членами домогосподарств.

Важливою складовою формування основи вибірки є упорядкування одиниць добору. Принципи упорядкування різні: за значеннями певної ознаки, за географічним критерієм, за абеткою або комбіновані. Правильно проведене упорядкування само собою забезпечує потрапляння у вибірку одиниць із різними значеннями ознак. Найпростішою основою вибірки є перелік одиниць генеральної сукупності, пронумерований від 1 до N . Це списки виборців, платників податків, домогосподарств, реєстри підприємств тощо. Перевагою такої основи є те, що:

- у процедуру добору включаються всі елементи генеральної сукупності;
- для обстеження можна дібрати будь-яку їх кількість, а саму вибірку здійснювати або за принципом лотереї, або з використанням генератора випадкових чисел.

Простими вважаються також набори звітів, анкет, карток, платіжних документів, оскільки кожен елемент такого набору визначає відповідну одиницю сукупності (торговельну угоду, пацієнта тощо). У цьому випадку одиниці основи є фізичними елементами, загальна їх кількість визначається прямим підсумком. Вони більш-менш однакові, їх можна змішати, як кульки в урні, й тим самим забезпечити випадковість добору.

У сукупностях, що складаються з «фізичних» одиниць, одиниця основи може репрезентувати як окремий елемент сукупності, так і певне їх угруповання (кластер, гніздо). Припустимо, що первинна сукупність має N одиниць, які утворюють M гнізд по N_j одиниць у кожному. Одиницею основи вибірки може бути одиниця сукупності, або гніздо. Відповідно формується вибіркова сукупність: у першому випадку вибирають n одиниць із загального числа N , у другому — m гнізд із загального числа гнізд M . Наприклад, рік можна розглядати як 364 дні або як 52 тижні по 7 днів. День є одиницею сукупності, тиждень — гніздом із 7 днів. При вивченні, скажімо, кількості опадів можна індивідуально вибрати 14 днів із загальної їх кількості (364), а можна вибрати 2 тижні по 7 днів.

Багато сукупностей не мають «природних» основ. Наприклад, не існує основи для вибірки дерев у лісі, виробів, що сходять з конвеєра при безперервному виробництві, туристів на Гаваях. У таких випадках використовують спеціально сконструйовані основи: карти (плани) міст чи сільських районів, площі, ділянки, смуги, лінії, точки. Наприклад, основою демографічного обстеження може бути нанесена на карту міста смуга, що перетинає кілька кварталів (перепис смуг), при обстеженні лісів — лінії (спосіб крейсування), при обстеженні стану посівів сільськогосподарських культур — ділянки.

Коли одиниці добору розміщені у просторі, основою вибірки виступають карти. До картографічних основ застосовують різноманітні схеми орієнтації на місцевості і способи добору одиниць спостереження, які мінімізують відстані пересування й забезпечують репрезентативність вибірки.

Традиційно карти і плани міст використовують в економіко-демографічних обстеженнях. Наприклад, при формуванні вибіркової сукупності домогосподарств у сільській місцевості карта/схема адміністративного району слугує основою упорядкування вибірових одиниць за принципом «географічного серпантину» (див. підрозділ 8.3).

Карти логічно виступають як основа вибірки в обстеженнях сільських територій і лісових угідь. Територію, що підлягає обстеженню, на карті поділяють на окремі ділянки, наприклад, площею в один гектар. При роботі з ділянками вирізняють два типи вибіркових одиниць:

- сама ділянка виступає одночасно як вибірка одиниця і як одиниця спостереження;
- ділянка розглядається як гніздо вибіркових одиниць, нанесених на карту у вигляді точок чи ліній.

Добір ділянок як одиниць основи виявився ефективним у сільськогосподарських, біологічних, інших обстеженнях. Дібрані ділянки обстежуються в натурі. Коли конструюється двоступенева основа вибірки, ділянка виступає як проміжна одиниця добору (гніздо), на другому ступені одиницями добору є точки або лінії всередині ділянок.

Якщо «природні» ділянки не мають достатньої кількості одиниць спостереження, то здійснюють вирівнювання ділянок за принципом рівності площ або за кількістю одиниць спостереження. Перший принцип застосовують при обстеженні урожаїв, у деяких екологічних обстеженнях тощо, другий — при обстеженні ферм, домогосподарств. Рис. 5.3. унаочнює обидва типи формування вибіркових одиниць.

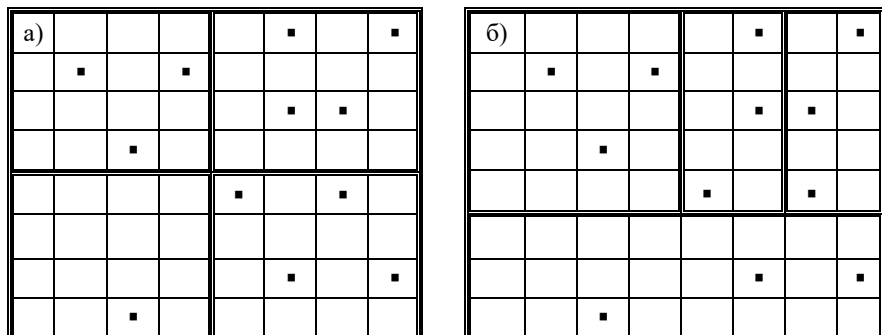


Рис. 5.3. Формування вибіркових одиниць першого ступеню (гнізд) за принципом: а) однакової площі; б) рівної кількості одиниць обстеження

Останнім часом посилюється роль аерокосмічних зображень як джерела інформації про стан посівів (зрідженість, забур'яненість, патологічні зміни, пошкодження шкідливими організмами), вплив

на посіви несприятливих погодних умов та потенційну урожайність сільськогосподарських культур. Аерокосмічні зображення виступають як різновиди карт, і при формуванні територіальної вибірки застосовують такі самі способи добору — за точками або за лініями, залежно від того, який із цих способів зручніший.

Особливим видом основи вибірки є статистична послідовність, наприклад, часові ряди, значення ознаки в яких відображають неперервні процеси і розміщені послідовно. Вибірковому спостереженню підлягають не всі значення послідовності, а відібрані через певні проміжки часу.

У багатоступеневій вибірці на кожному ступені формується «своя» основа. Як приклад можна навести вибіркове обстеження сільськогосподарської діяльності населення в сільській місцевості. Одиницею спостереження є домогосподарство, у володінні та користуванні якого є земельні ділянки, отримані відповідно до чинного законодавства.

Генеральну сукупність домогосподарств поділяють (стратифікують) за регіонами (областями та адміністративними районами), у межах кожної страти проводиться двоступенева процедура формування вибірки: на першому ступені добирають територіальні одиниці, на другому у межах територіальних одиниць добирають домогосподарства. Добір як територіальних одиниць, так і домогосподарств здійснюють пропорційно площі землі, яка перебуває у володінні та користуванні домогосподарств. Інформаційною базою обстеження є дані по господарського обліку.

Очевидно, в процесі формування багатоступеневої вибірки (зі своєю основою на кожному ступені добору) необхідно забезпечити взаємозв'язок між різними основами вибірки.

Важливим аспектом формування вибірки є визначення джерел інформації для формування основи вибірки. Інформаційна база основи вибірки має задовольняти таким вимогам, як: актуальність, повнота даних щодо одиниць досліджуваної сукупності, доступність даних, уніфікованість їх за змістом, одиницями вимірювання, формою подання. Безперечно, всі одиниці генеральної сукупності повинні мати необхідні для формування вибірки цільові ознаки; задля підвищення якості основи вибірки інформаційна база має бути актуалізована (оновлена) безпосередньо перед початком процедур з формування вибірки.

У практиці соціально-економічних досліджень багато сукупностей можна представити не однією, а низкою альтернативних

основ для вибірки. Скажімо, при обстеженні якості надання комунальних послуг жителям міста основою вибірки може бути список сімей або список житлових будинків; при обстеженні потенційної урожайності сільськогосподарських культур можливі такі основи, як списки господарств, карти полів, прямі лінії, що перетинають район, ділянки, точки. Одні з цих основ скінченні (списки господарств, ділянки), інші — нескінченні (прямі лінії, точки). Наукове обґрунтування та правильний вибір основи — одна з важливих передумов забезпечення репрезентативності результатів вибіркового спостереження.

При виборі основи вибірки необхідно перевірити, наскільки вона повна, чи не застаріла, чи не містить подвійного рахунку, чи відповідає меті дослідження, порівняти можливі витрати на обстеження, оцінити ефективність вибірових одиниць.

Припустімо, що для вибіркового оцінювання загальної суми значень ознаки Y є дві різні основи: g і k , які містять відповідно N_g і N_k одиниць. Параметри сукупності подано в табл. 5.1.

Таблиця 5.1

ПОРІВНЯННЯ ВІДНОСНОЇ ЕФЕКТИВНОСТІ ОСНОВ ВИБІРКИ

Основа	Кількість одиниць у сукупності	Значення параметрів		
		Середня	Дисперсія	Коефіцієнт варіації
g	N_g	\bar{y}_g	S_g^2	V_g
k	N_k	\bar{y}_k	S_k^2	V_k

Для оцінювання Y можна сформувати вибірку, дібравши n_g одиниць з N_g або n_k одиниць з N_k . Яка з цих основ вибірки більш ефективна? Щоб відповісти на це питання, слід порівняти:

- точність результатів, визначену на основі коефіцієнтів варіації, $V_\mu = \frac{V_x}{\sqrt{n}}$;

- витрати на обстеження $\frac{c_g}{c_k}$.

При порівнянні слід дотримуватися певних умов, зокрема щодо частки вибірки. Оскільки розміри одиниць у кожній основі можуть бути різними, використовують однакову в обох основах

частку вибірки, тобто $n_g / N_g = n_k / N_k$. Звідси надійність основи k відносно основи g визначається співвідношенням:

$$\gamma(k / g) = \frac{V_g^2 / n_g}{V_k^2 / n_k} = \frac{V_\mu^2(g)}{V_\mu^2(k)}.$$

Наприклад, для основи g обсяг вибірки $n_g = 12$, коефіцієнт варіації $V_g = 0,7$; для основи k , відповідно, $n_k = 20$, $V_k = 0,6$. Відносна надійність результатів, визначених за даними основи k , становить:

$$\gamma(k / g) = \frac{0,49/12}{0,36/20} = \frac{0,0408}{0,018} = 2,26,$$

тобто точність результатів, визначених за даними основи k , у 2,26 рази вища порівняно із результатами, визначеними за даними основи g .

Очевидно, різними будуть і витрати на обстеження вибірових одиниць різного типу чи розміру. Наприклад, вартість опитування фермерів щодо посівів ячменю менша за витрати з вимірювання площ під ячменем. У такому разі перевагу слід віддати вибірці ферм, оскільки за такою самою точністю даних вона дешевша. Нехай кошторис витрат на обстеження становить C , а вартість обстеження вибіркової одиниці в основах k і g , відповідно, c_k і c_g . Усю кошторисну суму можна витратити або на n_g одиниць основи g , або на n_k одиниць основи k . Звідси:

$$n_k = \frac{C}{c_k}, \quad n_g = \frac{C}{c_g}.$$

З урахуванням витрат на обстеження відносна ефективність основи визначається так:

$$\gamma^c(k / g) \approx \frac{V_\mu^2(g)c_g}{V_\mu^2(k)c_k} = \gamma(k / g) \frac{c_g}{c_k}.$$

Припустимо, у нашому прикладі $c_k = 3$, а $c_g = 2$. Тоді відносна ефективність основи k становить:

$$\gamma \frac{c_g}{c_k} = 2,26 \frac{2}{3} = 1,5,$$

тобто з урахуванням витрат на обстеження ефективність основи k у 1,5 раза вища.

Класичний приклад порівняння ефективності різних основ вибірки наводить Р. Джессен [Джессен, 1985: с. 81]. При дослідженні поширеності колорадського картопляного жука на сильно враженому полі було використано кілька основ вибірки: відрізки довжиною 48 футів в ряду посадки картоплі та різного розміру ділянки (гнізда). У табл. 5.2 наведено відносну ефективність гніздових основ порівняно з відрізками (1×48).

Таблиця 5.2

ВІДНОСНА ЕФЕКТИВНІСТЬ РІЗНИХ ОСНОВ ВИБІРКИ

Структура гнізда	1×48	2×2	4×4	6×6
Відносна ефективність, γ в %	100	51,9	19,0	9,3

У розглянутому прикладі ефективність гнізд будь-якого розміру нижча порівняно з окремими відрізками. Наприклад, у гнізді розміром (2×2) треба обстежити вдвічі більше відрізків, щоб отримати такий самий обсяг інформації, як при обстеженні відрізків, дібраних окремо. Це порівняння ефективності основ є зайвим доказом того, що при обстеженні природних сукупностей гніздові вибірки менш ефективні, ніж вибірки окремих одиниць.

Часом в основі вибірки трапляються невідповідності. У такому разі необхідно її уточнити або скоригувати. Найпростіший варіант — відсіяти зайві елементи вибірки на стадії проведення опитування. Наприклад, у маркетингових дослідженнях потенційним респондентам ставлять «запитання-фільтри», з'ясовують їхні соціально-демографічні характеристики, обізнаність щодо властивостей товару і досвід його використання. На підставі отриманих відповідей приймається рішення, скажімо, брати інтерв'ю у всіх респондентів чи лише у тих, хто ствердно відповів на запитання-фільтри.

Для коригування основи вибірки застосовують зважування, вирівнювання даних шляхом надання вагових коефіцієнтів. Наприклад, у вибірці 40 % чоловіків і 60% жінок, а в досліджуваній сукупності 46 % чоловіків і 54 % жінок, тоді кількість респондентів n , які обрали певну відповідь на запитання анкети, буде розраховуватися так:

$$n = \frac{46}{40} n_m + \frac{54}{60} n_w = 1,15 n_m + 0,90 n_w,$$

де \mathbb{Z}_m , \mathbb{Z}_w — кількість респондентів, чоловіків і жінок відповідно.

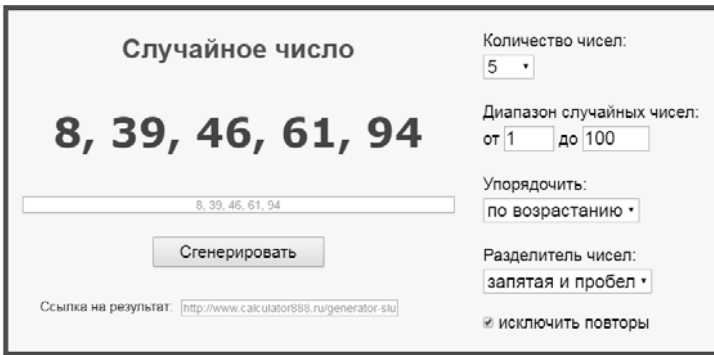
За необхідності підтримки двох і більше пропорцій розрахунок, ясна річ, ускладнюється. За відсутності статистичної інформації щодо детальної структури досліджуваної сукупності використовують спеціальні програмні продукти, які певною мірою мінімізують викривлення в структурі вибірки.

Формування та ведення основи вибірки передбачає максимально можливий рівень автоматизації робіт при складанні, коригуванні, актуалізації й архівуванні бази даних, використання сервісних процедур щодо роздруку, завантаження-вивантаження всієї чи певної частини основи вибірки, сортування за певними ознаками, отримання довідок щодо стану основи вибірки тощо. При цьому має бути забезпечена можливість автоматизованої процедури добору одиниць основи для обстеження.

Способи формування вибіркових сукупностей

Репрезентативність вибірки значною мірою залежить від того, в який спосіб формується вибіркова сукупність. У практиці вибіркових обстежень залежно від основи вибірки застосовують різні види випадкової вибірки як за організацією добору одиниць сукупності для обстеження, так і за кількістю ступенів добору. За організацією формування вибірки вирізняють способи, за яких добір відбувається без поділу генеральної сукупності на певні складові (простий випадковий добір), і способи, за яких такий поділ попередньо здійснюють (систематичний, стратифікований, кластерний). Розглянемо особливості кожного способу добору.

Простий випадковий добір (*simple random sampling*, або *SRS*) — це класичний спосіб формування вибіркової сукупності за принципом лотереї, який забезпечує усім одиницям генеральної сукупності рівні шанси бути дібраними для обстеження. Принцип випадковості добору реалізується за допомогою таблиці випадкових чисел (див. Додаток А) або генератора випадкових чисел (*random number, rnd*), який дає змогу отримати необхідну кількість випадкових чисел заданої розрядності й найширшого діапазону. Наприклад, в інтервалі від 1 до 100 необхідно отримати 5 випадкових чисел. За командою *Генерувати* маємо (рис. 5.4):



The image shows a web interface for a random number generator. At the top, it says "Случайное число" (Random number). Below that, the generated numbers are displayed in large, bold font: "8, 39, 46, 61, 94". To the right, there are several settings: "Количество чисел:" (Number of numbers) set to 5; "Диапазон случайных чисел:" (Range of random numbers) from 1 to 100; "Упорядочить:" (Sort) set to "по возрастанию" (by ascending); "Разделитель чисел:" (Number separator) set to "запятая и пробел" (comma and space); and a checked checkbox "исключить повторы" (exclude repeats). A "Сгенерировать" (Generate) button is located below the numbers. At the bottom left, there is a link "Ссылка на результат:" (Link to result) pointing to "http://www.calculator888.ru/generator-slu".

Рис. 5.4. Згенеровані випадкові числа

Згідно з отриманими випадковими числами зі сформованого списку основи вибірки добирають одиниці для обстеження, тобто формується вибіркова сукупність.

У великомасштабних соціально-економічних дослідженнях скласти повний список одиниць спостереження і здійснити вибіркове обстеження за класичною схемою простого випадкового добору практично неможливо. До того ж одиниці спостереження зазвичай розпорошені на великій території, що ускладнює організацію дослідження і збільшує витрати на його проведення. Тому на практиці застосовують різні модифікації випадкової вибірки. Кожен варіант модифікації, точно кажучи, не є випадковою вибіркою, проте організується в такий спосіб, який забезпечує максимальне наближення до умов випадкового добору і водночас дозволяє помітно підвищити економічну ефективність вибірки та розширити сферу її використання.

Коли основа вибірки впорядкована у певний ідентифікований спосіб, її можна поділити на стільки інтервалів, скільки необхідно дібрати одиниць для обстеження, і в кожному інтервалі визначити ті особливі «числові» одиниці, які формують вибірку. Інші модифікації випадкового добору пов'язані з урахуванням структури генеральної сукупності, попередньою її класифікацією за певними ознаками. Існує два основні види класифікації генеральної сукупності: стратифікація (районування) і кластеризація (серійно-гніздова вибірка). Поєднання різних способів добору урізноманітнюють клас модифікованих випадкових вибірок. Розглянемо їхні характерні особливості.

Найпоширенішим способом формування вибірки і найпростішим з точки зору організації є *систематичний добір* (*systematic sampling*). Сутність його полягає в доборі одиниць із впорядкованої основи вибірки (списки, реєстри, довідники, специфікації, тощо) через рівні інтервали (рівномірно розподілений добір). Крок добору h обчислюється діленням обсягу генеральної сукупності N на передбачений обсяг вибірки n :

$$h = N / n.$$

Якщо це дріб, округляють до найближчого цілого числа.

Перший випадковий номер — початкову одиницю добору Z_1 — визначають множенням кроку добору h на випадкове число rnd , яке набуває значень в інтервалі від 0 до 1. Отримати rnd можна за допомогою генератора випадкових дробових чисел (*random fractional numbers*). Отже:

$$Z_1 = h \times rnd.$$

Наступна одиниця добору залежить від початкової Z_1 і кроку інтервалу. Так, для частки вибірки $\frac{n}{N} = 0,05$ кроком добору є число $\frac{N}{n} = \frac{1}{0,05} = 20$, тобто у вибірку має потрапити кожна двадцять

та одиниця. Якщо випадкове число $rnd = 0,35$, то перша одиниця вибірки буде під номером $20 \times 0,35 = 7$, друга — під номером $(7 + 20) = 27$, третя — $(27 + 20) = 47$ і т.д.

Можлива ситуація, коли номер останньої одиниці, яку необхідно дібрати для обстеження, виявиться за межами обсягу генеральної сукупності (у списку відсутній). У такому разі список можна подовжити або за рахунок випадково дібраних одиниць, або за принципом «замкненого кола» — після одиниці з останнім номером розмістити одиниці з першими номерами.

На точність систематичної вибірки може істотно вплинути невдалий вибір початкової одиниці добору. Скажімо, формується вибірка квартир для обстеження з інтервалом 20. Усі будинки п'ятиповерхові, на кожному поверсі по 4 квартири. Якщо початковою одиницею добору взяти квартиру № 3, то до вибірки потраплять квартири, розташовані на першому поверсі: 23, 43, 63 і т.д., якщо взяти квартиру № 17, то до вибірки потраплять всі квартири, розташовані на п'ятому поверсі. Через небезпеку ви-

никнення систематичної похибки практикують використання кількох початкових одиниць, які почергово застосовують до різних інтервалів.

Систематична вибірка забезпечує рівномірний добір одиниць і пропорційне представництво їх усіх сегментів сукупності. Порівняно з простою випадковою систематична вибірка ефективніша, її простіше здійснити. Однак точність результатів обстеження залежить від наявності випадковості в розміщенні одиниць генеральній сукупності. При застосуванні систематичного добору в ряду, ранжируемому за кількісною ознакою, може виникнути систематична похибка репрезентативності. Коли первинний матеріал попередньо не впорядкований, тобто одиниці сукупності розміщені випадково щодо цільової ознаки, то систематичний добір можна розглядати як різновид простої випадкової вибірки.

Специфічною формою систематичного добору є поширена в аудиторській практиці монетарна вибірка, в якій елементом добору є не господарська операція, а відображена в обліку грошова сума. Тобто починаючи з випадково обраного числа, інтервал добору визначається не кількістю елементів сукупності (наприклад, кожен десятий документ), а вартістю елементів сукупності (наприклад, елементи, на які припадають кожні 100 тис. грн).

Систематичний добір використовують також у процесі поточного контролю якості технологічного процесу шляхом періодичного взяття виготовлених на той момент виробів. Поточний контроль дає змогу своєчасно виявити відхилення від встановленого режиму роботи й запобігти дефектам і браку продукції.

Особливим видом систематичного добору вважають *моментне* спостереження, суть якого полягає в періодичній фіксації стану досліджуваного явища. На відміну від інших вибірових спостережень моментне не має чіткого розмежування між генеральною і вибірковою сукупністю, воно є суцільним за охопленням одиниць сукупності і вибіровим за кількістю моментів спостереження.

Класичними прикладами моментної вибірки є спостереження за використанням виробничого устаткування чи робочого часу на підприємствах. Скажімо, в цеху, де працює 40 осіб, проводиться моментне спостереження використання робочого часу. Тривалість одного обходу — 15 хв., обходи проводяться кожні 30 хвилин. Якщо робоча зміна триває 8 годин (480 хв.), то 30-хвилинних від-

різків буде 16, максимально можливе число записів за зміну — $40 \times 16 = 640$, а час спостереження становитиме $640 \times 30 = 38400$ хв. Це і є генеральна сукупність моментного спостереження. Вибір моментів для спостереження організують за схемою систематичної вибірки, тобто через рівні проміжки (інтервали) часу. Кількість записів визначається при проектуванні спостереження за процесом і залежить від варіації досліджуваного показника, який досліджується, від особливостей виробничого процесу, розташування робочих місць чи розміщення устаткування, тривалості періоду спостереження.

Систематичний добір можна здійснити без попереднього списку, спираючись на територіальне розміщення одиниць у вигляді географічних карт (див. підрозділ 8. 3).

Стратифікований спосіб формування вибірки (*stratified sampling*) на відміну від простого випадкового чи систематичного, які проводяться в цілому по генеральній сукупності, передбачає попередній поділ основи вибірки N на m страт (груп, районів, пластів) і незалежний добір певної кількості одиниць у кожній страті n_j :

$$\left\{ \begin{array}{l} N_1 \rightarrow n_1 \\ N_2 \rightarrow n_2 \\ \dots\dots\dots \\ N_m \rightarrow n_m \end{array} \right\}$$

Обсяг стратифікованої вибірки n визначається як сума вибірок по окремих стратах, тобто $n = \sum_j^m n_j$. При цьому розміщення вибі-

рки обсягом n за стратами n_j здійснюється або пропорційно чисельності окремих складових генеральної сукупності N_j , або пропорційно варіації ознаки в цих складових (див. розділ 7).

У статистичному аспекті стратифікація відповідає виокремленню такої кількості і таких статистично однорідних груп, усередині яких варіація цільових ознак була б меншою, ніж відмінності між групами.

Слід зауважити, що організація стратифікованої вибірки має сенс лише в неоднорідних сукупностях. Страти створюються так, щоб кожна з них за можливості представляла однорідні групи одиниць. Оскільки в однорідних групах варіація ознак значно

менша, то за однакового обсягу вибірки n стратифікація забезпечує вищу точність результатів обстеження, ніж простий випадковий чи систематичний добір. Проблема полягає в тому, як саме здійснити стратифікацію, щоб досягти якнайбільшого ефекту. Задавою стосовно стратифікації має бути ознака, за якою одиниці сукупності істотно різняться і яка взаємопов'язана з цільовою ознакою. При обстеженнях населення ознакою стратифікації можуть бути типи поселень, соціальні, вікові чи освітні групи; при обстеженні підприємств — види економічної діяльності, форми власності, розміри, тощо. У більшості великомасштабних опитувань громадської думки за національними вибірками первинну стратифікацію здійснюють за географічною локалізацією респондентів (звідси й синонімічна назва цього принципу — районування).

Процедури стратифікації різняться за:

- кількістю ознак стратифікації (їх може бути дві і більше);
- способом добору одиниць з кожної страти (пропорційний, непропорційний).

Кластерний спосіб формування вибірки (*cluster sampling*) ґрунтується на випадковому доборі з генеральної сукупності не окремих одиниць, а певних цілісних утворень (серій, гнізд, кластерів), які мають бути максимально ідентичними щодо заданих параметрів. У більшості випадків генеральна сукупність має чітке природне розчленування на кластери, які складаються з одиниць, об'єднаних або територіально (райони, селища), або організаційно (школи, лікарні, колективи). Можливі й «штучно» утворені кластери, наприклад, кластери регіонів за інвестиційною привабливістю. Вибіркова сукупність кластерів формується за схемами систематичного або простого випадкового добору. У межах кластерів, залежно від їхньої чисельності та однорідності, здійснюють або суцільне обстеження (однорівнева кластерна вибірка), або добирають одиниці для обстеження, застосовуючи власне випадковий чи систематичний спосіб формування вибірки (дворівнева кластерна вибірка).

Кластерна вибірка порівняно з іншими має певні переваги організаційного характеру, які особливо відчутні при формуванні кластерів за територіальною ознакою: добір невеликої кількості територіальних одиниць (населених пунктів, районів, житлових кварталів тощо) істотно скорочує терміни проведення і знижує вартість обстеження.

Ефективність кластерного способу формування вибірки залежить від того, які саме кластери дібрано для обстеження. Елементи всередині кластера мають бути якомога різноманітнішими (гетерогенними), а самі кластери — схожими між собою. Особливе значення має співвідношення кількості включених до вибірки кластерів і обсягу вибірки з кожного кластера. За будь-яких умов краще збільшувати кількість кластерів і зменшувати обсяги вибірки з кожного кластера. Проте конкретні співвідношення між параметрами вибірки залежать від організаційних можливостей та наявних ресурсів (див. розділ 8).

Основна відмінність кластерної вибірки від стратифікованої полягає в тому, що в стратифікованій вибірці всі підгрупи (страти) використовуються для подальшого добору, а в кластерній вибірці обстежуються лише дібрані кластери. Цільові установки цих способів добору також різні. Мета стратифікованої вибірки — підвищувати точність вибіркових оцінок, мета кластерної вибірки — підвищити ефективність вибірки, зменшивши витрати на її проведення.

Якісне різноманіття соціально-економічних явищ, складність і багатогранність взаємозв'язків потребують поєднання різних способів формування вибірки, наприклад, комбінація кластерного і власне випадкового добору, стратифікованого і систематичного тощо. Комбіновані вибірки (*combined sample*) здатні забезпечити високу репрезентативність результатів з найменшими трудовими і грошовими витратами на організацію та проведення досліджень.

Вибірки складнішого дизайну різняться за кількістю ступенів добору. При одноступеневій вибірці (*one-stage sample*) усі випадково дібрані одиниці чи кластери підлягають обстеженню. Процес формування багатоступеневої вибірки (*multistage sample*) відбувається у кілька послідовних етапів. Ступенів може бути два, три й більше. На кожному з них своя, відмінна від інших основа вибірки. Відповідно поділяються й одиниці вибірки (репрезентації): першого ступеня, другого і т. д.

При проектуванні багатоступеневих вибірок вирізняють первинні, проміжні й кінцеві одиниці вибірки, які співвідносяться між собою за принципом «матрьошки». Первинними одиницями вибірки найчастіше виявляються територіально-адміністративні одиниці, кінцевими — ті одиниці, сукупність яких є безпосереднім об'єктом обстеження, проміжними — організації, в яких кон-

центруються кінцеві одиниці обстеження. Скажімо, при вивченні професійної орієнтації молоді кінцевими одиницями вибірки є школяри, а проміжними — школи.

Вибір одиниць для безпосереднього обстеження здійснюється на останньому ступені формування вибіркової сукупності. Частка її відносно до генеральної сукупності залежить від часток вибірки на всіх ступенях. Якщо припустити, що до вибірки потрапив один із десяти районів ($d_1 = 0,10$), у цих районах дібране кожне п'яте селище ($d_2 = 0,20$), а у дібраних селищах обстежується 4% домогосподарств ($d_3 = 0,04$), то частка вибіркової сукупності в генеральній становить:

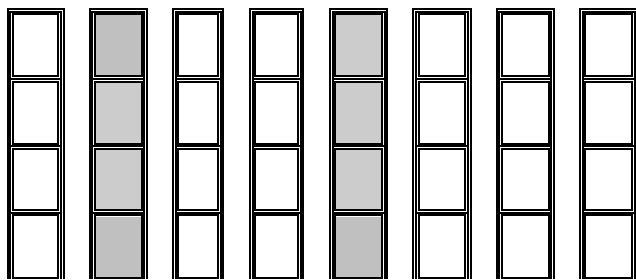
$$d = d_1 d_2 d_3 = 0,1 \cdot 0,2 \cdot 0,04 = 0,0008 ,$$

тобто обстеженню підлягає 0,08 % домогосподарств.

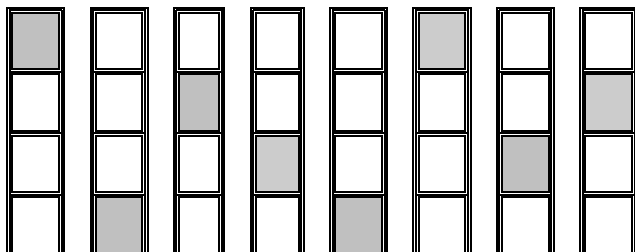
Багатоступенева вибірка значно зменшує витрати на обстеження й порівняно з іншими вибірками ефективніша. Широке практичне застосування такої форми організації вибіркового спостереження зумовлене передусім її гнучкістю, що спрощує весь процес організації спостереження.

На рис. 5.5 схематично подано особливості формування кластерної, стратифікованої і двоступеневої вибірок. Сукупність обсягом 24 одиниці поділена на 8 сегментів (гнізд) по 4 одиниці. Обсяг вибірки — 8 одиниць. При кластерній вибірці добирають 2 гнізда по 4 одиниці, при стратифікованій — по одній одиниці з кожного гнізда, при двоступеневому доборі спершу добираються чотири гнізда, а потім — у кожному відібраному гнізді — по 2 одиниці.

а) Кластерна вибірка



б) Стратифікована вибірка



в) Двоступенева вибірка

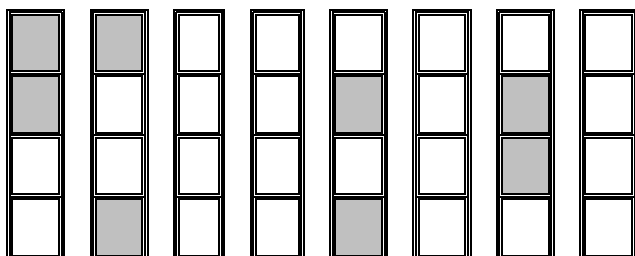


Рис. 5.5. Схеми організації різних способів добору

У ситуації, коли обстежують сукупність за двома і більше ознаками, які різняться варіацією, а необхідну точність їхніх оцінок забезпечує різний обсяг вибірки, ефективною виявляється *багатофазна вибірка (multiphase sample)*. Сутність її полягає в тому, що для різних ознак формуються вибіркові сукупності різного обсягу: з генеральної сукупності утворюється первинна вибірка, з первинної — підвибірка і т. д. На відміну від багатоступеневої вибірки, коли основа вибірки на різних ступенях добору інша, а безпосередньому обстеженню підлягають одиниці, дібрані на останньому ступені, у багатофазній вибірці одні й ті самі одиниці добору зберігаються на всіх фазах обстеження, при цьому на кожній наступній фазі обсяг підвибірки зменшується, а програма обстеження G {кількість запитань} — розширяється. Наприклад, під час дослідження ставлення студентів факультету до організації навчального процесу у 1-й фазі досліджують 25 % від загального їх числа за короткою програмою ($n_1 = 100$; $G\{3\}$), у 2-й фазі — 20 % за ширшою програмою, яка містить ще три до-

даткові запитання, у 3-й фазі — 5 % за повною програмою із десяти запитань (рис. 5.6).

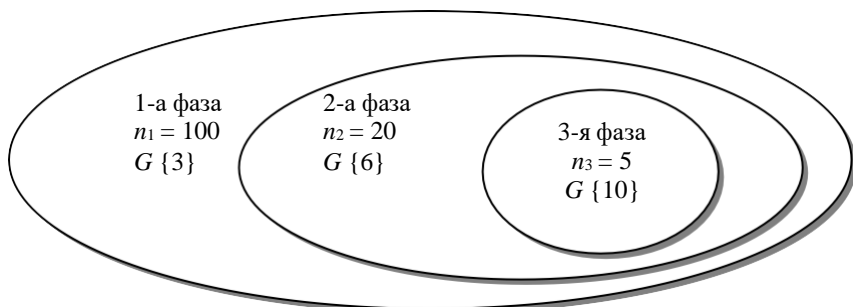


Рис. 5.6. Макет формування трифазної вибірки

Безперечною перевагою багатофазної вибірки є те, що даними, отриманими на першій фазі, можна користуватися як додатковою інформацією на другій фазі, даними другої фази — як додатковою інформацією на третій фазі і т.д., що підвищує точність результатів вибіркового обстеження. При організації багатофазної вибірки можливі комбінації різних способів і видів вибірки (систематичної, стратифікованої, кластерної, багатоступеневої). Похибки репрезентативності розраховують окремо для кожного фази.

На різних ступенях добору, а часто й на всіх, вихідний (природний) список проміжних одиниць попередньо стратифікується. Добір гнізд здійснюється зі страт, а отримана в такий спосіб вибірка є вже «гібридом» кластерного і стратифікованого методів добору. Добір гнізд із попередньо сформованих страт найкраще робити за допомогою процедур власне випадкової або систематичної вибірок. Мінімальна кількість дібраних гнізд із кожної страти не менше двох.

Іноді виникає потреба здійснити дві (або більше) незалежні між собою, рівновеликі вибірки з однієї і тієї сукупності, застосовуючи один і той самий спосіб добору. Такі вибірки дістали назву *взаємопроникненних*. Порівняння результатів обстеження незалежних вибірок дає підстави для оцінювання надійності вибіркових даних: чим менші відхилення між оцінювання різних вибірок, тим надійнішими є результати вибіркового обстеження. Взаємопроникненні вибірки зручні для порівняння результатів

дослідження за окремими географічними районами. До такого типу вибірок вдаються також, коли треба за короткий час отримати попередні підсумки обстеження. Водночас при організації вибіркових обстежень слід зважати на те, що взаємопроникненні вибірки порівняно з іншими способами добору потребують значно більших трудових і грошових витрат.

Застосування того чи іншого способу формування вибіркової сукупності залежить від мети вибіркового обстеження, можливостей його організації та проведення. Найпоширенішими є комбіновані вибірки, які поєднують різні способи добору і здатні забезпечити високу репрезентативність результатів з найменшими трудовими і грошовими витратами на організацію та проведення досліджень.

У зв'язку з необхідністю моніторингу швидкоплинних соціально-економічних явищ і процесів поширюється практика повторних вибіркових обстежень. При їх організації обсяг і склад кожної наступної вибірки може залишатися незмінним, а може змінюватися. Залежно від того, чи є спільна частина у повторних вибірок, і яка саме частина одиниць, обстежених у попередній вибірці, включається в наступну, вирізняють три варіанти організації повторних вибіркових обстежень: повне заміщення, ротаційна вибірка і панельне обстеження (рис. 5.7).

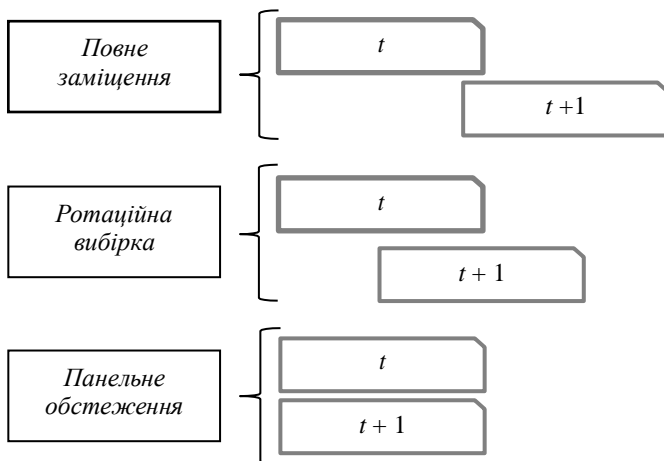


Рис. 5.7. Варіанти організації повторних вибірок

Сформована шляхом повного заміщення повторна вибірка не вимагає, щоб склад її був однаковим у різні моменти; важливо, щоб ці вибірки належали до однієї генеральної сукупності. Повторну вибірку з частковою заміною сукупності називають *ротаційною (rotational sampling)*. Ступінь ротації може бути різною, найчастіше коливається в межах від 0,25 до 0,75.

Панельне обстеження в класичному вигляді не передбачає заміщення одиниць попередньої вибірки. Вибіркова панель — це постійно функціонуюча вибіркова сукупність, яка регулярно надає інформацію за відповідною програмою обстеження. Вибіркові панелі утворюються для порівняльних досліджень. Власні панелі мають великі маркетингові служби, спеціалізовані маркетингові й консалтингові фірми, науково-дослідницькі організації.

Коли немає чіткого уявлення про об'єкт дослідження, а дослідники не можуть висунути жодних наукових гіпотез, проводять пробні (пошукові, пілотажні) обстеження, обсяг вибірки яких становить від 5 до 10% обсягу основної вибірки. Мета пробного обстеження — уточнити програмно-методологічні питання обстеження, сформулювати гіпотези, перевірити придатність та надійність дослідницького інструменту і за потреби внести необхідні корективи та зміни в організацію основного дослідження.

При формуванні основи вибірки типовою проблемою є наявність пропусків у масивах даних (неповне охоплення генеральної сукупності). Існує багато способів відтворення пропусків (див. підрозділ 10.1); важливо застосувати той спосіб, який здатен зберегти особливості генеральної сукупності як основи вибірки і забезпечити репрезентативність вибірових даних.

Застосування не випадкового (спрямованого) добору

Поряд з розглянутими способами формування вибірових сукупностей, які ґрунтуються на випадковому доборі, в практиці вибірових обстежень застосовують методики, які не передбачають дотримання принципу випадковості добору. Це можуть бути пошукові дослідження, а також ситуації, в яких неможливо сформувати основу вибірки. Добір одиниць для обстеження в таких вибірках відбувається переважно за суб'єктивними критеріями: доступність, цілеспрямованість, типовість, рівне представництво

тощо. Згідно з цими критеріями, включення одиниці сукупності до вибірки розглядається як результат рішення або дослідника, або респондента. Конкретні форми не випадкового добору визначають:

- доступність об'єкта вивчення, тобто готовність дослідника включати до вибірки тих, хто бажає брати участь в дослідженні й охоче надає інформацію;
- цілеспрямованість добору, тобто прагнення дослідника формувати вибірку з тих одиниць, яким властиві певні ознаки.

Основний недолік таких процедур формування вибірки пов'язаний з відсутністю статистичних методів, які б дозволяли зробити обґрунтовані висновки щодо досліджуваної сукупності. Оцінювання точності і валідності результатів обстеження залишається справою суб'єктивного судження і досвіду.

Найбільш регламентованим різновидом не випадкового, спрямованого добору є *квотований (quota sampling)*, методика проведення його спрямована на відтворення в структурі вибірки певних пропорцій генеральної сукупності за обраними показниками у вигляді квот. У формуванні квотованої вибірки засадовими є не перелік (список) одиниць спостереження, а відомості про найважливіші пропорції генеральної сукупності. Кількість одиниць, що належить обстежити за кожною квотою, визначається з урахуванням попередньої інформації про структуру генеральної сукупності. У такий спосіб намагаються досягти адекватності вибіркової і генеральної сукупностей.

За визначеними пропорціями, виходячи із загального обсягу вибіркової сукупності, встановлюють відповідні квоти, які доводять до відома тих, хто має брати інтерв'ю. Наприклад, при вивченні громадської думки інтерв'юєр має опитати десять менеджерів-чоловіків віком 30–40 років, п'ять домогосподарок віком 40–50 років і т. ін. У який спосіб «заповнити квоти» респондентами з потрібним поєднанням цих ознак, інтерв'юєр вирішує самостійно, не покладаючись на випадок.

Цілеспрямоване забезпечення подібності вибіркової і генеральної сукупностей за визначеними параметрами досягається зазвичай за допомогою вибірки істотно меншого обсягу, ніж при використанні випадкового добору. Мінімальний обсяг вибірки у поєднанні з певною свободою при доборі конкретних одиниць дослідження є істотними організаційними перевагами квотованого методу, що зумовлює міру його поширеності переважно в соціологічних і маркетингових дослідженнях. А оскільки зменшен-

ня обсягу вибірки поєднується зі зменшенням витрат і скороченням терміну проведення дослідження, то це підсилює переваги зазначеного способу добору. З іншого боку, оскільки добір квотами за своєю природою не випадковий, існує велика ймовірність суб'єктивного підходу до формування вибірки, що не гарантує репрезентативності вибіркового оцінок.

Проблема ефективного застосування методу квот має два аспекти. Перший стосується апріорної інформації про об'єкт дослідження. Застосування методу квот передбачає наявність актуалізованої (повної і новітньої) інформації про структуру генеральної сукупності. Другий аспект пов'язаний із множиною ознак, за якими генеральна сукупність структурована і на які спирається квотований добір. Здавалося б, чим більше пропорцій генеральної сукупності враховується у квотований вибірці, тим вищою має бути її репрезентативність. Проте чим більше ознак в умовах визначеного обсягу вибірки, тим більше обмежень, яких необхідно дотримуватися при формуванні вибірки. А чим більше обмежень, тим менше допустимих рішень, тим важче підібрати таку структуру вибірки, яка б забезпечувала репрезентативність результатів.

Зазвичай при формуванні квотованої вибірки використовують дві-три незалежні ознаки, збільшення їх кількості значно ускладнює реалізацію квотованого добору. Припустимо, соціолог вирішив використати 4 ключові ознаки (стать, вік, освіту і вид діяльності) й забезпечити відповідність вибірки генеральній сукупності за співвідношенням осіб, яким властиві всі чотири ознаки. Якщо перша ознака має 2 градації, друга — 5, третя — 3, а четверта — 7, то кількість груп, які соціолог вирішив взяти під контроль, становитиме $(2 \times 5 \times 3 \times 7) = 210$. Щоб забезпечити таке саме, співвідношення груп у вибірці, як у генеральній сукупності, обсяг вибірки має бути досить великим, але в такому разі втрачається основна перевага квотованої вибірки — зменшення її обсягу.

Через неможливість відтворити у вибірковій сукупності всі характеристики генеральної сукупності, обмежуються ключовими, які презентують основні соціальні групи, види діяльності, території тощо. Якщо створена модель вибірки відповідає структурі генеральної сукупності за ключовими ознаками, можна припустити, що така вибірка певною мірою відтворить пропорції генеральної сукупності і за іншими ознаками, які з різних причин не були враховані.

У соціологічних і маркетингових дослідженнях модель квотованої вибірки найчастіше створюється на основі пропорцій, які відповідають основним соціально-демографічним характеристикам генеральної сукупності: стать, вік, освіта, професія, соціальний статус, місце проживання. Це зумовлено тим, що: а) ці ознаки мають ключовий характер і впливають на більшість характеристик сукупності; б) інформацію щодо цих ознак можна отримати зі статистичних джерел.

До переваг квотованої вибірки належать доступність та зручність її структури, короткі терміни формування, незначні витрати на збирання інформації. До недоліків — неможливість точного вимірювання відхилень, зумовлених процесом цілеспрямованого добору одиниць.

Основна проблема «класичного типу» квотованої вибірки полягає в тому, що вона формується особою інтерв'юєром. Відсутність контролю за добром респондентів призводить до появи систематичних похибок, розмір яких не піддається вимірюванню. Крім того, у квотованій вибірці такого типу важко визначити кількість відмов від відповідей. З наведених причин квотований спосіб формування вибірки зазнав критики, що, власне, посприяло його вдосконаленню за рахунок більшої «рандомізації». Так, до квотованого добору почали включати елементи випадкової вибірки, зокрема, пошук респондентів інтерв'юєри ведуть не самостійно, а працюють зі списками осіб, обираючи з них носіїв потрібного поєднання ключових ознак. При організації багатоступеневої вибірки квотований добір застосовують лише на останньому ступені її формування. Якщо на попередніх ступенях дотримуватися принципу випадковості добору, то така стратегія забезпечить самозважування вибірки за найважливішими ознаками, завдяки чому ризик, пов'язаний із квотованим добром, зменшиться.

Особливою сферою застосування цілеспрямованого добору є формування фокус-груп і груп експертів, коли необхідно спиратися як на соціально-демографічні характеристики, так і на показники наукової і практичної діяльності.

Різновидом цілеспрямованої вибірки є спосіб *снігового клубка* (*snowball sampling*). Цей спосіб найчастіше застосовують при вивченні відносно нечисленних груп респондентів, які перебувають поза сферою інституціонального контролю, наприклад, спеціалістів рідкісних професій, представників елітних груп, групи коле-

кціонерів тощо. Власне, це техніка пошуку і добору респондентів з певним поєднанням властивостей в умовах, коли важко окреслити межі генеральної сукупності. Особливість методу полягає в тому, що, за винятком першого кроку, добір кожного чергового респондента відбувається за рекомендацією респондентів, включених до вибірки на попередньому кроці. Незважаючи на те, що перша група респондентів добирається випадково, сформована у такий спосіб вибірка не репрезентативна, оскільки соціально-демографічних характеристик осіб, яких рекомендує респондент, схожих між собою більше, ніж це могло бути за випадкового добору.

У ситуації, коли дослідник не має змоги забезпечити відповідність вибіркової сукупності генеральній за всіма ознаками, що характеризують одиницю спостереження, застосовують вибірку доступних одиниць. Це один із видів стихійного добору, найширше його застосовують у соціально-психологічних дослідженнях, у кінодокументалістиці, журналістиці, тобто у сферах, де дотримання строгих вимог не є критично обов'язковим. У соціологічних дослідженнях вибіркова сукупність також може формуватися з максимально доступних (потенційних) респондентів, яким власне й належить рішення про включення до вибірки. Наприклад, дослідження «маятникової міграції» населення. Опитування респондентів відбувається в електропоїздах, автобусах тощо. Для інтерв'юєра важлива лише належність респондента до «мігрантів» та його згода відповісти на запитання інтерв'юєра (респондент сам долучає себе до вибірки).

Усі види стихійного опитування мають істотні недоліки. По-перше, для вихідної сукупності не можна побудувати репрезентативну вибірку. По-друге, неможливо теоретично визначити похибку, що виникає за такої схеми добору. Вказані недоліки не означають, втім, що стихійний добір недоцільний. У деяких випадках за його допомогою можна дізнатися, наприклад, про варіанти поведінки телеаудиторії, ставлення до політичних партій, їхніх лідерів тощо. Стихійний добір часто буває корисним у пробних та пілотних дослідженнях. Незамінний він і в дослідженнях інтимних аспектів життя людей. Принцип випадкового добору для дослідження таких об'єктів використати практично неможливо, бо частка відмов від бесіди набагато перевищує граничну норму. Різні способи стихійного добору мають неоднакову міру ефективності. Однак чим більшою є міра випадковості в доборі

одиниць дослідження, тим точнішим, за інших однакових умов, є його результат. Водночас недбало здійснений імовірнісний добір, по суті, може перетворитися на стихійний.

Обираючи спосіб формування вибірки, треба враховувати призначення дослідження, співвідношення ціна-якість, часові рамки, прийнятність помилок. Попри на відсутність статистичних методів, які б давали змогу узагальнити результати, отримані у не випадкових вибірках, і поширити їх на генеральну сукупність, дослідники визнають доцільним використання такого типу вибірок у пошукових (тестових та пілотних) дослідженнях, а також при формуванні експериментальних і контрольних груп, члени яких дали згоду на участь в експерименті, груп експертів, в опитуваннях громадської думки тощо. В разі, коли результати вибіркового обстеження будуть узагальнюватися або коли дослідження має підтвердити чи спростувати певну гіпотезу, перевагу надають імовірнісній, репрезентативній вибірці.

Вправи та запитання для самоконтролю

1. Поясніть сутність вибіркового спостереження. Які його переваги перед іншими видами спостереження?
2. Проведено обстеження:
 - кожного десятого сільськогосподарського підприємства зі 150, які проводять зрошення земель за рахунок власних коштів, з метою вивчення ефективності використання зрошувальних площ;
 - агрофірми «Степова» з метою вивчення резервів підвищення ефективності зрошення саме в цьому господарстві.Яке з обстежень є вибіркоvim?
3. Що означає репрезентативність вибірки? За яких умов вибірка буде репрезентативною?
4. Чому принцип випадковості добору є визначальним при формуванні вибіркової сукупності?
5. Із наведених вибірових обстежень визначте, дані яких обстежень містять систематичну похибку реєстрації, а яких — систематичну похибку репрезентативності:
 - при формуванні вибіркової сукупності робітників з метою вивчення продуктивності праці наперед були виключені робітники зі стажем менше 1 року;
 - при обстеженні стану тваринництва у фермерських господарствах через недбалість обліковців у деяких господарствах не повністю було враховано молодняк худоби.

6. Обґрунтуйте основу вибірки, вкажіть вибіркочну одиницю і одиницю спостереження при обстеженні:

- цін на пальне;
- ефективності телереклами;
- витрат домогосподарств на продукти харчування;
- якості закупленої у фермерів молочної продукції.

7. Назвіть способи формування вибіркової сукупності.

8. Як здійснюється власне випадковий добір? У чому різниця між повторним та неповторним способами добору?

9. Чим відрізняється систематичний добір від власне випадкового?

10. Поясніть сутність стратифікації і порядок формування стратифікованої вибірки.

11. У чому полягає особливість кластерної вибірки?

12. Поясніть сфери використання стратифікованої і кластерної вибірки.

13. Поясніть особливості та сфери використання багатоступеневої і багатофазної вибірки.

14. Вкажіть вид вибірки в таких обстеженнях:

– при обстеженні бюджету часу студентів університету було дібрано кожного десятого бакалавра і кожного двадцятого магістра;

– при перевірці податкових декларацій за минулий рік у районній податковій службі добиралася кожна десята декларація;

– при обстеженні терміну користування короткостроковим кредитом у банківських установах добирався кожен п'ятий банк із загального їх списку, а потім у дібраних банках — кожен сотий рахунок.

15. Проектується вибіркоче обстеження домогосподарств у сільській місцевості з метою визначення розміру земельних ділянок і джерел формування доходів. Можливі такі способи добору:

– добирається кожен п'ятий населений пункт, і в кожному з них — десяте домогосподарство;

– добирається кожне десяте домогосподарство із загального списку домогосподарств.

Для кожного способу добору вкажіть вибіркочну одиницю і одиницю спостереження. Який спосіб добору доцільно застосувати?

16. При обстеженні площ, засіяних озимим ячменем, можливі дві основи вибірки: ділянки і ферми. Розмір і варіація цих основ характеризується даними:

Характеристики	Ферми (g)	Ділянки (k)
Кількість одиниць	200	450
Загальна площа під ячменем, га	4800	4800
Середня врожайність ячменю, ц/га	56	63
Дисперсія врожайності	529	361

Порівняйте надійність основ вибірки, а також їхню ефективність за умови, що витрати часу на опитування одного фермера становлять 10 хв., а витрати часу на пошук і вимірювання ділянки — 25 хв.

17. За яких умов використовують не випадкові вибірки?

18. Поясніть сутність квотованого добору й можливості його рандомізації.

19. Поясніть особливості формування вибірки способом «снігового клубка».

20. Чому стихійний добір не вважають випадковим? Якою є сфера його застосування?

Рекомендована література

1. Вибіркове спостереження: Термінологічний словник / Укл.: О.О. Васечко, О.І. Черняк, Є.М. Жуйкова, О.В. Гончар та ін. — К. : Держкомстат України, 2004. — 140 с.

2. *Гончар О.В.* Забезпечення якості статистичної інформації: методологія та організація : Монографія. — К. : ВПД «Формат», 2011. — 372 с.

3. *Джессен Р.* Методы статистических обследований. — Москва: Финансы и статистика, 1985. — 478 с.

4. *Єріна А.М.* Організація вибірових обстежень : Навчальний посібник. — К. : КНЕУ, 2003. — 127 с.

5. *Черняк О.І.* Техніка вибірових досліджень. — К. : МІВВЦ, 2001. — 248 с.

РОЗДІЛ 6

ТОЧНІСТЬ ВИБІРКОВИХ ОЦІНОК

Функції і властивості вибіркового оцінок
Стандартна похибка як міра точності вибіркового оцінок
Дизайн вибірки та дизайн-ефект
Визначення мінімально достатнього обсягу вибірки

Після вивчення цього розділу студент має

- *знати*: вимоги щодо надійності результатів вибірки, співвідношення надійності і точності вибіркового оцінок; вимірники точності; основні методологічні підходи до розроблення дизайну вибірки, який забезпечує необхідну точність вибірки;
- *уміти*: оцінювати точність результатів вибірки і визначати довірчі межі значень цільових показників; задля забезпечення необхідної точності результатів обстеження розробляти план (дизайн) вибірки і визначати мінімально достатній обсяг вибірки з урахуванням кошторису дослідження.

Функції і властивості вибіркового оцінок

Головне завдання вибіркового обстеження — за результатами спостережень оцінити невідомі параметри генеральної сукупності з відповідною точністю та достатньою надійністю. За умови випадковості добору теоретичним підґрунтям оцінювання є закон великих чисел, згідно з яким за необмеженого збільшення числа спостережень $n \rightarrow \infty$ вибіркові оцінки наближаються (сходяться за ймовірністю) до відповідних параметрів генеральної сукупності.

Нехай випадкова величина X підпорядкована закону розподілу $F(X; Z_1, \dots, Z_k)$, де Z — параметри розподілу (частка, середня, дисперсія тощо), числові значення яких невідомі й підлягають оцінюванню. Правило, за яким на підставі даних вибірки обсягом n розраховується оцінка невідомого параметра генеральної сукупності Z , називають функцією вибірки $z = f(x_1, x_2, \dots, x_n)$. Будь-яка ви-

біркова оцінка z , яка містить інформацію про невідомий параметр генеральної сукупності Z , залежить від числових значень ознаки у дібраних одиниць сукупності x_i ($i = 1, 2, 3, \dots, n$), обсягу вибірки n і закону розподілу досліджуваної випадкової величини x .

Щоб вибіркові оцінки були максимально наближеними до справжніх значень оцінюваних параметрів, а втрата інформації, необхідної для прийняття статистичних рішень, була мінімальною, вони мають бути надійними й задовольняти певним вимогам. Основними властивостями статистичних оцінок є: спроможність, незміщеність і статистична ефективність.

Вибіркова оцінка z невідомого параметра Z є *спроможною*, коли вона підпорядковується закону великих чисел, тобто зі збільшенням числа спостережень з імовірністю як завгодно близькою до одиниці можна стверджувати, що відхилення між параметром Z і його оцінкою z не перевищить будь-яку скільки завгодно малу величину $\alpha > 0$:

$$P\{|Z - z| < \alpha\} \rightarrow 1.$$

Вимога спроможності необхідна, щоб оцінка мала практичне значення, інакше збільшення обсягу вибірки не буде «наближати до істини». Середні величини, частки розподілу, дисперсії є спроможними оцінками відповідних параметрів генеральної сукупності. З іншого боку, спроможність — це асимптотична властивість, яка може проявитися лише в сукупностях великого обсягу. Можна запропонувати кілька спроможних оцінок одного і того самого параметра. Це свідчить про те, що однієї спроможності для повної характеристики надійності оцінок недостатньо.

Друга властивість вибіркової оцінки — *незміщеність*. Оцінка z вважається незміщеною, якщо для будь-якого обсягу вибірки n її математичне сподівання дорівнює дійсному значенню параметра Z :

$$E(z) = Z.$$

Якщо ця рівність не виконується, то оцінка z може або завищувати, або занижувати значення параметра Z , тобто йдеться про систематичну (одного знака) похибку в оцінці параметра Z . Вимога незміщеності оцінки (за дотримання вимоги спроможності) особливо важлива у разі малої кількості спостережень.

Незміщену оцінку z , яка має найменшу дисперсію серед усіх можливих незміщених оцінок параметра Z , визначених для вибі-

рок такого самого обсягу, називають *ефективною* оцінкою. Поняття ефективності безпосередньо пов'язане зі стандартною похибкою відповідної функції оцінки. За наявності кількох оцінок найефективнішою буде та з них, у якої стандартна похибка мінімальна.

Незміщеність та ефективність розглядаються як вирішальні властивості при виборі оцінок: незміщеність свідчить про *надійність* оцінки, ефективність — про її *точність*. Оцінка може мати високу надійність, але низьку точність і навпаки. Сутність і співвідношення понять точності й надійності оцінок розглянемо на умовному прикладі влучення в ціль пострілами з гвинтівки (рис. 6.1).

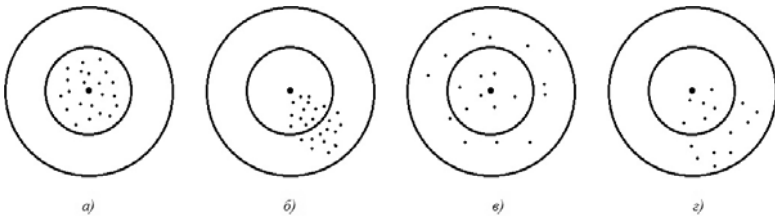


Рис. 6.1. Точність та надійність: спільні риси і відмінності

Як видно, гвинтівка **а**) б'є прямо по мішені, тоді як постріли інших гвинтівок помітно відхиляються від неї. Проте характер відхилень різний. Постріли з гвинтівки **б**) концентруються у нижньому правому куті мішені, завдяки компактності їх можна вважати точними, але вони зсунені, ненадійні. Сенсу від такої точності немає, оскільки постріли не можуть уразити ціль. Попадання з гвинтівки **в**) хоча й надійні (незміщені), але надто розкидані, що свідчить про низьку їх точність. Нестійкі, неточні результати гвинтівки **в**) свідчать, що одна лише надійність може бути марною властивістю вибірки. Нарешті, гвинтівка **г**) має і низьку надійність, і низьку точність.

Для статистичного оцінювання параметрів генеральної сукупності бажано використовувати оцінки, які одночасно задовольняють вимоги спроможності, незміщеності й ефективності. На практиці задовольнити одночасно всім трьом вимогам щодо вибірових оцінок не завжди вдається. Насамперед це стосується спільного виконання вимог незміщеності й ефективності оцінок. У такому разі найчастіше перевагу віддають зміщеним, але точнішим оцінкам перед незміщеними, що мають низьку точність.

Прикладна статистика використовує два типи оцінок параметрів генеральної сукупності — точкові та інтервальні. *Точкова оцінка* — це конкретне числове значення функції $z = f(x_1, x_2, \dots, x_n)$, де x_i — результати n спостережень над кількісною ознакою випадкової величини x . Саме точкова оцінка визнається найкращим наближенням до справжнього значення оцінюваного параметра генеральної сукупності. *Інтервальною оцінкою* називають інтервал, який з певною ймовірністю покриває справжнє значення параметра генеральної сукупності.

Для знаходження «найкращих» точкових оцінок невідомих параметрів генеральної сукупності застосовують різні методи, серед них поширеніші такі: метод моментів, метод найменших квадратів і метод максимальної правдоподібності. Використання методу моментів ґрунтується на припущенні, що початкові й центральні моменти вибіркової сукупності є достатніми оцінками відповідних моментів генеральної сукупності. Оцінки методом моментів зазвичай спроможні, порівняно прості в розрахунках, але іноді малоефективні.

В основу оцінювання методом найменших квадратів покладено умову мінімізації суми квадратів відхилень вибірових значень функції від теоретичних. Цей метод широко використовують в аналізі явищ, що змінюються в часі, та в регресійному аналізі, проте він забезпечує ефективність оцінок лише стосовно нормального розподілу.

Суть методу максимальної правдоподібності полягає в тому, що із можливих оцінок невідомого параметра Z обирається та, якій відповідає найбільша імовірність, тобто те значення, яке обертає функцію в максимум. Оцінки максимальної правдоподібності, як правило, спроможні, асимптотично ефективні та незначно зміщені.

Як доведено математичною статистикою, вибіркова середня є незмщеною, спроможною й ефективною оцінкою генеральної середньої, а вибіркова дисперсія — зміщеною оцінкою генеральної дисперсії. Математичне сподівання вибіркової дисперсії $E(s^2)$ дещо менше дисперсії генеральної сукупності:

$$E(s^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{(n-1)}{n} \cdot \sigma^2,$$

де σ^2 і s^2 — дисперсія генеральної сукупності і її вибіркова оцінка відповідно.

Для коригування зміщення вибіркової дисперсії використовують коефіцієнт $\frac{n}{n-1}$ (поправку Бесселя). виправлену в такий спосіб вибірку дисперсію розглядають як незміщену оцінку дисперсії генеральної сукупності:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}.$$

У малих вибірках, чисельність яких варіює в межах від 5 до 30, поправочний коефіцієнт відхиляється від одиниці, тому завжди потрібно користуватися незміщеною оцінкою дисперсії. Зі збільшенням n поправка прямує до одиниці, а при $n > 50$ різниця між зміщеною і незміщеною оцінками практично відсутня і поправкою можна знехтувати.

Точкові оцінки навіть у тих ситуаціях, коли вони спроможні (наближаються до дійсного значення параметра при збільшенні n), незміщені (у середньому збігаються зі значенням параметра Z) і ефективні (з найменшою дисперсією), — це випадкові величини, яким притаманний неконтрольований розкид і певна статистична похибка, а статистичні висновки у зв'язку з цим мають імовірнісний характер.

З огляду на випадкову природу точкової оцінки z виникає необхідність визначення діапазону допустимих значень параметра Z для заданої ймовірності. Такий діапазон називають *довірчим* інтервалом, а його межі — *довірчими* межами. Найчастіше довірчий інтервал обирають симетричним до параметра Z , тобто:

$$F(z - \Delta \leq Z \leq z + \Delta) = 1 - \alpha.$$

Ширина довірчого інтервалу свідчить про точність оцінки вибірки: чим вужчий інтервал для ймовірності $(1 - \alpha)$, тим точніша вибірка оцінка. Ймовірність $(1 - \alpha)$ характеризує ступінь надійності оцінки. Якщо $(1 - \alpha) \cdot 100 = 95\%$, значить, у 95 випадках зі 100 невідомий параметр покривається довірчим інтервалом. Звідси ймовірність помилкового висновку (рівень істотності α) становить $100 - 95 = 5\%$.

Вибір рівня істотності залежить від рівня відповідальності за прийняті рішення. В одних випадках необхідна досить висока гарантія точності статистичного висновку, ймовірність помилко-

вого рішення не має перевищувати 0,1 %, тобто $\alpha \leq 0,001$, в інших прийнятним може бути рівень істотності $\alpha \leq 0,10$.

У прикладній статистиці застосовують прямі і непрямі методи оцінювання. *Пряме оцінювання* здійснюють на основі прямої оцінної функції за даними, отриманими в процесі конкретного вибіркового обстеження за результатами безпосереднього вимірювання. Основою *непрямого оцінювання* є непряма оцінна функція, при побудові якої передбачається використання даних, отриманих як за результатами безпосереднього вимірювання (на основі конкретного обстеження), так і з використанням додаткової (зовнішньої) інформації (наприклад, даних аналогічних обстежень по інших територіях, за інші періоди, дані інших обстежень тощо). У сучасній статистичній практиці цей метод став ефективним інструментом розв'язання проблеми підвищення надійності оцінок показників на різних рівнях агрегації даних (див. підрозділ 10.3).

Стандартна похибка як міра точності вибірових оцінок

Одна з головних характеристик надійності результатів вибірових обстежень — точність вибірових оцінок, наближеність їх до реальних (невдомих) значень параметрів генеральної сукупності. Точність інформації — категорія відносна, вона пов'язана з такими категоріями як надійність та достовірність. Не можна очікувати абсолютної точності вибірових оцінок, але необхідно забезпечити такий рівень їхньої точності, який гарантуватиме користувачеві прийняття ефективних управлінських рішень.

Мірою точності вибірової оцінки є похибка вибірки, яка показує, наскільки оцінка z може відрізнитися від справжнього значення параметра Z при наявному дизайні вибірки та обсязі вибірової сукупності. Очевидно, чим менше відхилення $|z - Z|$, тим точніша оцінка. Точність будь-якої вибірової оцінки залежить від двох факторів: методу оцінювання і способу формування вибірової сукупності. У теорії вибіркового методу всі доведення ґрунтуються на простій випадковій вибірці, де немає обмежень на дію механізму закону великих чисел (саме цей закон визначає логічну конструкцію вибіркового методу).

Припустімо, що з певної генеральної сукупності здійснено безліч вибірок однакового обсягу, і показники окремих вибірок утворили ряди можливих значень вибірових середніх, часток, дисперсії тощо. Для кожної вибірки характерна своя похибка. Якщо побудувати ряди розподілу похибок вибірки для кожного показника, то за характером розподілу похибок ряди будуть різнитися (симетричні, асиметричні), але для всіх розподілів характерною буде концентрація похибок навколо центру розподілу.

Як уже зазначалося, розподіли похибок підпорядковуються певним законам, описуються відповідними функціями, властивості яких слугують підґрунтям для статистичних висновків щодо ймовірностей того чи іншого рівня похибки та можливих значень невідомих параметрів генеральної сукупності. Розподіл вибірових середніх наближається до нормального закону, незалежно від того, нормально чи ні розподілена сукупність, з якої сформовані вибірки; дисперсія вибірових середніх у n разів менша від дисперсії ознаки в генеральній сукупності, тобто:

$$\mu^2 = \sigma_0^2 / n.$$

Середнє квадратичне відхилення вибірових оцінок від значення параметра в генеральній сукупності називають *стандартною похибкою вибірки* μ (*standard error*). Оскільки на практиці генеральна дисперсія ознаки σ_0^2 невідома, в розрахунках можна використати вибірову незміщену оцінку дисперсії: для повторної вибірки $\sigma^2 \frac{n}{n-1}$, для безповторної $\sigma^2 \frac{N-n}{N-1}$. Отже, формули стандартної похибки такі:

для повторної вибірки

$$\mu = \sqrt{\frac{\sigma^2}{n-1}},$$

для безповторної вибірки

$$\mu = \sqrt{\frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)}.$$

Коригувальний множник $\left(\frac{N-n}{N-1} \right)$ називають «поправкою на скінченність сукупності», а обернена до нього величина розглядається як міра відносної надійності схеми безповторного добору. Оскільки це відношення перевищує одиницю для будь-якого

обсягу вибірки, окрім $n = 1$, можна сказати, що результати безповторної вибірки завжди надійніші порівняно із повторною:

$$\left| \frac{N-h}{n} \right| \geq 1.$$

Практика вибірових обстежень спирається саме на схеми безповторного добору, які забезпечують високу точність вибірових оцінок при меншому обсязі вибірки. Щодо практичного використання наведених формул слід урахувати таке:

а) у сукупностях, поділених за бінарною ознакою, яка набуває лише два значення (**1** — наявність і **0** — відсутність певної властивості), основним параметром оцінювання є частка одиниць сукупності з певною властивістю; дисперсія частки, як відомо, становить:

$$\sigma^2 = p(1-p) = pq,$$

де p і q — частки вибіркової сукупності з відповідними значеннями ознаки, сума цих часток ($p + q$) = 1;

б) у великих за обсягом сукупностях (30 і більше одиниць) поправка $\frac{n}{n-1}$ не вносить істотних змін у розрахунки, а тому береться до уваги лише в малих вибірках;

в) коригувальний множник для безповторної вибірки $\sqrt{\frac{N-n}{N-1}} \approx \sqrt{1-\frac{n}{N}}$ при малих величинах $\frac{n}{N}$ наближається до 1, що видно з даних табл. 6.1. Отже, у випадках, коли частка обстеженої сукупності $\frac{n}{N} < 5\%$, розрахунок похибки вибірки можна виконувати за формулою для повторної вибірки; коригувальний множник доцільно застосовувати при $\frac{n}{N} \geq 5\%$.

Таблиця 6.1

ЗНАЧЕННЯ КОРИГУВАЛЬНОГО МНОЖНИКА ПРИ РІЗНИХ ЧАСТКАХ ВИБІРКИ

$\frac{n}{N}$	0,01	0,05	0,10	0,20
$\sqrt{\frac{N-n}{N-1}} \approx \sqrt{1-\frac{n}{N}}$	0,995	0,975	0,949	0,894

Формули розрахунку стандартних похибок вибірки для інших параметрів сукупності (моди, медіани, моментів розподілу 3-го і 4-го порядків, коефіцієнтів кореляції і регресії) наведено у Додатках Б1-Б5.

У практичних розрахунках зазвичай використовують не стандартну, а *граничну похибку* Δ , яка є добутком стандартної похибки μ на довірче число t :

$$\Delta = t\mu.$$

Значення t залежить від імовірності, з якою гарантують результат оцінювання. Зв'язок між ними виражається інтегралом розподілу ймовірностей Лапласа-Гаусса $F(t)$. Змінна t приймає додатні і від'ємні значення; математичне сподівання $E(t) = 0$, дисперсія — $\sigma_0^2(t) = 1$. Із симетричності нормального розподілу випливає, що $F(0) = 0,50$; $1 - F(t) = F(-t)$. Зазначені властивості стають очевидними при геометричній інтерпретації функції.

Гранична похибка вибірки $\Delta = t\mu$ — це максимально можлива похибка для взятої ймовірності $F(x)$. Довірче число t вказує, як співвідносяться гранична й стандартна похибки. Як бачимо з рис. 6.2, з імовірністю 0,683 гранична похибка не вийде за межі стандартної $\Delta \pm 1\mu$, з імовірністю 0,954 вона не перевищить $\pm 2\mu$, з імовірністю 0,997 — $\pm 3\mu$. На практиці найчастіше використовують імовірність 0,954 (на рис. 6.2 незаштрихована частина площини).

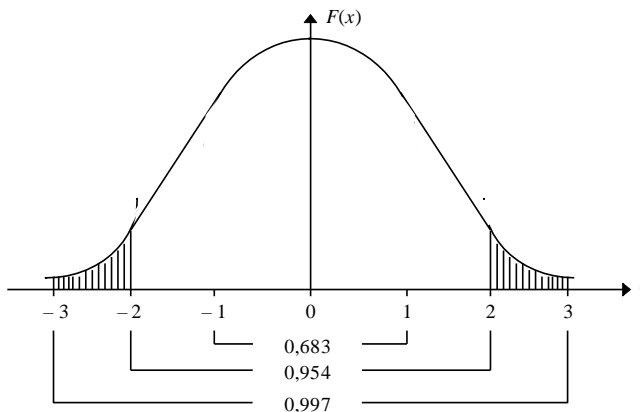


Рис. 6.2. Співвідношення ймовірностей та граничної похибки

З урахуванням викладених формул граничних похибок середньої та частки записують так:

	Повторна вибірка	Безповторна вибірка
Для середньої	$\Delta_{\bar{x}} = t \sqrt{\frac{\sigma^2}{n}}$	$\Delta_{\bar{x}} = t \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$
Для частки	$\Delta_p = t \sqrt{\frac{pq}{n}}$	$\Delta_p = t \sqrt{\frac{pq}{n} \left(1 - \frac{n}{N}\right)}$

Як видно з формул, розмір граничної похибки залежить від:

- варіації ознаки σ^2 ,
- обсягу вибірки n ,
- частки вибірки в генеральній сукупності $\frac{n}{N}$,
- узятого рівня довіри, якому відповідає квантиль t .

Чим більша варіація ознаки в генеральній сукупності, тим більша в середньому похибка вибірки. Залежність похибки від обсягу вибіркової сукупності обернено пропорційна. Щоб зменшити похибку вибірки вдвічі, обсяг останньої має зрости в 4 рази. При безповторному доборі похибка буде тим меншою, чим більшою буде частка обстеженої сукупності $\frac{n}{N}$. Очевидно, при суцільному

спостереженні похибка репрезентативності відсутня ($\Delta = 0$).

Оцінюючи невідомий параметр генеральної сукупності Z , важливо визначити не лише точкову його оцінку, а й ступінь надійності (достовірності) статистичного висновку щодо його рівня, тобто вказати інтервал оцінних значень із наперед заданою ймовірністю. Центром такого інтервалу, як правило, є точкова оцінка z , а ширина інтервалу залежить від стандартної похибки μ та прийнятої ймовірності статистичного висновку. Отже, межі довірчого інтервалу визначають на основі точкової оцінки та граничної похибки вибірки $\Delta = t\mu$, де μ — стандартна похибка вибірки:

для середньої

$$\bar{x} - t\mu \leq \bar{x}_0 \leq \bar{x} + t\mu ;$$

для частки

$$p - t\mu \leq p_0 \leq p + t\mu .$$

При визначенні довірчих меж рівень довіри беруть не менше 80 %, найбільш уживаний — 95 %, тобто рівень істотності α не перевищує $100 - 95 = 5$ %. Оскільки нормальний розподіл симетричний, то $t_\alpha = -t_{1-\alpha}$, а $|t|_{1-\alpha} = t_{1-\frac{\alpha}{2}}$.

Модулі значень коефіцієнта довіри t для нормального розподілу наведено в табл. 6.2.

Таблиця 6.2

ЗНАЧЕННЯ $t_{1-\frac{\alpha}{2}}$ СТАНДАРТНОГО НОРМАЛЬНОГО РОЗПОДІЛУ

Рівень довіри, %	80	85	90	95	98	99
$t_{1-\frac{\alpha}{2}}$	1,28	1,44	1,64	1,96	2,33	2,58

У практиці вибірових спостережень та в науково-дослідницькій роботі часто доводиться користуватися невеликими за обсягом вибірками, чисельність яких не перевищує 30 одиниць (агрономічні й зоотехнічні досліди, перевірка якості продукції, пов'язана з руйнуванням зразків, тощо). У статистиці вони дістали назву малих вибірок. Оскільки вибіркова дисперсія в малих вибірках ($n < 30$) зміщена, то в розрахунках стандартних похибок вибірки використовують незміщену оцінку дисперсії $s^2 = \sigma^2 \frac{n}{n-1}$.

Засади теорії малих вибірок розробив англійський математик-статистик В. Госсет (псевдонім Стьюдент). Він винайшов t -розподіл, який не залежить від параметрів генеральної сукупності, а залежить лише від одного параметра — обсягу вибірки і дозволяє робити імовірнісні висновки для малої кількості спостережень. При $n \rightarrow \infty$ розподіл Стьюдента наближається до нормального. У Додатку В 1 наведено квантили t -розподілу Стьюдента для ймовірності 0,95 і числа ступенів свободи $df = n - 1$ ¹. При $n > 30$ квантили нормального і t -розподілу практично збігаються.

Розглянемо методику вибіркового оцінювання середньої та частки на прикладі обстеження швидкості розрахунків з кредиторами однієї фірми.

¹ df — degree of freedom (ступені свободи).

Приклад. З метою визначення швидкості розрахунків з кредитором однієї фірми була здійснена вибірка 100 платіжних документів, за якими середній термін перерахунку і отримання грошей виявився 27,2 дня з дисперсією 30,25. Кожен десятий платіжний документ був із порушенням терміну розрахунків.

Визначимо середній термін проходження всіх платежів упродовж року і частку платежів з порушенням термінів розрахунку з імовірністю 0,954 ($t = 2$).

Гранична похибка для середнього терміну проходження платежів:

$$\Delta_x = 2 \sqrt{\frac{30,25}{100}} = 2 \cdot 0,55 = 1,1 \text{ дня.}$$

Це дає підстави стверджувати з імовірністю 0,954, що середній термін проходження платежів упродовж року щонайменше 26,1 дня і не перевищує 28,3 дня.

$$27,2 - 1,1 \leq \bar{x} \leq 27,2 + 1,1.$$

Перш ніж визначити граничну похибку для частки платежів з порушенням терміну розрахунків, необхідно обчислити дисперсію частки:

$$s^2 = 0,1(1 - 0,1) = 0,09.$$

Гранична похибка частки платежів з порушенням терміну оплати:

$$\Delta_p = 2 \sqrt{\frac{0,09}{100}} = 0,06, \text{ або } 6,0 \%.$$

Що стосується інтервалу можливих значень частки платежів з порушенням терміну оплати в генеральній сукупності, то межі його досить широкі: від 4,0 до 16,0 %:

$$10 - 6,0 \leq p \leq 10 \pm 6,0.$$

У статистичному аналізі часто виникає необхідність порівняти похибки вибірки різних ознак або однієї самої ознаки в різних сукупностях.

Такі порівняння виконують за допомогою *відносної стандартної похибки* V_μ , яка показує, на скільки процентів вибір-

кова оцінка може відхилятися від параметра генеральної сукупності:

$$V_{\mu} = \frac{\mu_x}{\bar{x}} \cdot 100.$$

Її розмір можна визначити також за коефіцієнтом варіації ознаки V_x :

для повторної вибірки,

$$V_{\mu} = 100 \frac{V_x}{\sqrt{n}}$$

для безповторної вибірки

$$V_{\mu} = 100 \frac{V_x}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}.$$

Так, у нашому прикладі стандартна похибка для середнього терміну проходження платежів становить $\mu = 0,55$, а відносна похибка:

$$V_{\mu} = \frac{0,55}{27,2} 100 = 0,02, \text{ або } 2\%.$$

Аналогічний результат дає розрахунок відносної похибки за коефіцієнтом варіації:

$$V_x = \frac{\sqrt{30,25}}{27,2} 100 = 20,2\%.$$

$$V_{\mu} = \frac{20,2}{\sqrt{100}} = 2,0\%.$$

Похибку вибірки для частки Δp також слід порівнювати з величиною частки p , адже одна й та сама похибка $\Delta p = 2\%$ для $p = 80\%$ є малою, для $p = 40\%$ — допустимою, для $p = 10\%$ — великою. Відносну похибку частки обчислюють за формулою:

$$V_{\mu} = \frac{\mu_p}{p} = \frac{\sqrt{\frac{pq}{n}}}{p} = \sqrt{\frac{q}{np}}.$$

У нашому прикладі стандартна похибка частки платежів з порушенням терміну оплати становить $\mu = 3,0\%$, а відносна:

$$V_{\mu} = \frac{3,0}{10,0} = \sqrt{\frac{0,9}{100 \cdot 0,1}} = 0,3, \text{ або } 30\%,$$

що значно більше порівняно із похибкою для середнього терміну проходження платежів (2,0 %).

Отже, відносну стандартну похибку можна використати для порівняння вибірових оцінок різних ознак. Іноді використовують граничну відносну похибку, яка враховує ймовірність статистичного висновку $V_{\Delta} = tV_{\mu}$.

Розглянутий порядок оцінювання точності вибірових оцінок математично доведений і статистично обґрунтований для простого випадкового добору. Коли випадкова вибірка формується в інший спосіб, застосовують модифіковані вимірники точності/надійності вибірових оцінок з урахуванням особливостей відповідного способу добору (див. підрозділ 6.3).

У світовій практиці вибірових обстежень величину V_{μ} розглядають як рівень придатності вибірових оцінок для аналізу. Так, достатньо високою визнається точність на рівні $V_{\mu} \leq 3\%$, а на рівні $V_{\mu} > 20\%$ — малоприйнятною. У табл. 6.3 наведено класи точності вибірових даних.

Таблиця 6.3

КЛАСИ ТОЧНОСТІ ВИБІРКОВИХ ДАНИХ

Клас	Стандартна відносна похибка вибірки, %	Рівень точності
1	2 і менше	Висока точність
2	3–5	Підвищена точність
3	6–10	Середня точність
4	11–20	Низька точність
5	понад 20	Малоприйнятна точність

Для класифікації надійності передбачається шість класів: практично достовірні дані (ймовірність вища 99,7 %); з малим ризиком (95 %); із середнім ризиком (80 %); з підвищеним ризиком (60 %); азартні (менше 60 %), а також невизначені дані, коли ймовірність невідома. На вибір класу надійності впливають розмір можливих втрат від похибок розрахунку, можливість компенсації відхилень, а також фактори суб'єктивного порядку, зокрема рівень відповідальності осіб, які прийматимуть рішення на підставі даних вибірового обстеження.

Дизайн вибірки та ефект дизайну

Дизайн вибірки — це стратегія формування вибіркової сукупності, яка містить: визначення генеральної сукупності та основи вибірки; критеріїв стратифікації та (або) кластеризації; структури багатоступеневої вибірки і методів добору на різних її ступенях; схеми ротації вибірових одиниць у повторних обстеженнях. Фундаментом вибору цієї стратегії є: мета дослідження, умови його фінансування, особливості генеральної сукупності. При розробленні дизайну вибірки обов'язково враховують обмеження та умови, визначені технічним завданням. Сучасні вибірові обстеження, насамперед державні, здебільшого мають складний дизайн, який дозволяє вирішувати проблеми оптимізації вартості обстеження і забезпечує вимоги щодо надійності його результатів.

Числовим виразом статистичної ефективності певного дизайну D слугує запропонований Л. Кішем [Kish, 1995] індикатор надійності вибірових оцінок — дизайн-ефект (*design effect* — *deff*). Визначається *deff* співвідношенням квадратів стандартних похибок: у чисельнику квадрат похибки за реального дизайну вибірки $\mu^2(D)$, у знаменнику — квадрат похибки за умови, що вибірка сформована за принципом простого випадкового добору, μ^2 :

$$deff = \frac{\mu^2(D)}{\mu^2}.$$

Індикатор *deff* вимірює вплив дизайну вибірки на точність вибірових оцінок. Якщо *deff* < 1, то ефективність плану (дизайну) вибірки D вища порівняно з простим випадковим добром, при *deff* > 1, навпаки, ефективність вибірового плану менша. Зауважимо, що дизайн-ефект розраховується для кожного показника окремо.

Очевидним є взаємозв'язок між стандартними похибками випадкового добору і певного дизайну:

$$\mu^2(D) = deff \mu^2 = deff \frac{s^2}{n}.$$

Слід пам'ятати, що дисперсія для частки визначається за формулою:

$$s^2 = pq = p(1 - p).$$

З урахуванням дизайн-ефекту вибірки довірчі межі для середньої та частки можна записати так:

$$\bar{x} \pm t\sqrt{deff} \sqrt{\frac{s^2}{n}}; p \pm t\sqrt{deff} \sqrt{\frac{pq}{n}}.$$

Поряд з прямим співвідношенням квадратів похибок (*deff*) при порівнянні точності і надійності результатів за різними планами вибірки використовують також обернену величину $\gamma = \frac{1}{deff}$.

Очевидно, що при $\gamma > 1$ маємо вигреш у точності, а при $\gamma < 1$, навпаки, втрату точності.

Розрахунок стандартної похибки вибірки адаптується до специфіки способів формування вибірових сукупностей. Так, систематичний добір застосовують до даних, розміщення яких має елементи і випадковості, і певної упорядкованості. Похибки при систематичному доборі зумовлені не способом добору, а наявністю випадковості в розміщенні одиниць сукупності. Коли одиниці сукупності розміщені за випадковою схемою, наприклад, за абеткою назв окремих одиниць, тоді систематичний добір рівнозначний випадковому неповторному, і розрахунок похибки вибірки здійснюється відповідно до формулою неповторної вибірки:

$$\mu_{syst}^2 = \frac{s^2}{n} \cdot \left(\frac{N-n}{N-1} \right) \approx \frac{s^2}{n} \cdot \left(1 - \frac{n}{N} \right).$$

Практично дизайн-ефект систематичної вибірки визначається часткою одиниць генеральної сукупності, яка добирається для обстеження, тобто $deff_{syst} = 1 - \frac{n}{N}$.

Так, для 4%-ої вибірки $deff = 0,96$, для 10%-ої — 0,90. Процедура систематичного добору легко поширюється на випадки, коли одиниці сукупності різні за розміром і добір здійснюється з імовірністю, пропорційною їхньому розміру (див. підрозділ 8.1).

В інших модифікаціях випадкової вибірки (стратифікація, кластеризація) при розрахунку похибки вибірки спираються на відому теорему декомпозиції варіації, згідно з якою загальна дисперсія s^2 поділяється на окремі компоненти, наприклад, на міжгрупову s_B^2 і внутрішньогрупову (середню з групових дисперсій) s_W^2 :

$$s^2 = s_B^2 + s_W^2.$$

Розрахунок вибірових оцінок цих дисперсій здійснюють за схемою дисперсійного аналізу (табл. 6.4).

Таблиця 6.4

СХЕМА ДИСПЕРСІЙНОГО АНАЛІЗУ

Джерело варіації	Сума квадратів відхилень, S	Число ступенів свободи, df	Оцінка дисперсії (середній квадрат відхилень), S / df
Між групами	$\sum_1^m (\bar{y}_j - \bar{y})^2 n_j$	$m-1$	s_B^2
Всередині груп (внутрішня)	$\sum_1^m \sum_1^{n_j} (y_{ij} - \bar{y}_j)^2$	$n-m$	s_W^2
Загальна	$\sum_1^m \sum_1^{n_j} (y_{ij} - \bar{y})^2$	$n-1$	s^2

Якщо в розрахунках похибки вибірки враховувати не всю варіацію, а лише окремі її компоненти, тобто елімінувати вплив або міжгрупової, або внутрішньої варіації, то значення похибки зменшиться, а отже, підвищиться точність результатів вибіркового обстеження.

Так, при визначенні стандартної похибки стратифікованої вибірки враховується лише внутрішня варіація, мірою якої є оцінка дисперсії s_W^2 . Якщо сформовані групи об'єднують «схожі» елементи, а групові середні величини помітно різняться, варіація ознаки в групах буде значно меншою, ніж для сукупності загалом. У такому разі $s_W^2 < s^2$, а отже, похибка стратифікованої вибірки порівняно із простою випадковою чи систематичною буде меншою:

$$\mu_{strat}^2 = \frac{s_W^2}{n},$$

де $n = \sum_1^m n_j$.

Отже, точність стратифікованої вибірки тим вища, чим істотніші міжгрупові розбіжності. Ознака розшарування і ознака, характеристики якої оцінюються, співвідносяться між собою як причина й наслідок, зв'язок між ними за своєю природою кореляційний.

За правилом декомпозиції варіації середню з групових дисперсій можна визначити як різницю $s_w^2 = s^2 - s_B^2$, або як скориговану на щільність кореляційного зв'язку загальну дисперсію $s_w^2 = s^2(1 - r^2)$, де r — коефіцієнт кореляції:

$$\mu_{strat}^2 = \frac{s^2(1 - r^2)}{n}.$$

Як бачимо, розшарування сукупності зменшує похибку вибірки на частку $(1 - r^2)$. Чим щільніший зв'язок між ознаками, тим помітніше зменшення похибки. При $r^2 = 0,50$ похибка вибірки зменшується вдвічі, при $r^2 = 0,66$ втричі.

Дизайн-ефект стратифікованої вибірки порівняно із простою випадковою при обстеженні одного і того самого об'єкта менший за одиницю:

$$deff_{strat} = \frac{\mu_{strat}^2}{\mu^2} \approx \frac{\sum_{j=1}^m s_j^2}{s^2 n} < 1.$$

При обчисленні стандартної похибки кластерної вибірки, в якій кожен з дібраних для обстеження кластер розглядається як одне ціле, враховується лише міжгрупова (міжкластерна) варіація:

$$\mu_{clast}^2 = \frac{s_B^2}{m},$$

де s_B^2 — міжкластерна дисперсія;

m — чисельність дібраних для обстеження кластерів.

Міжкластерна дисперсія — частина загальної дисперсії $s_B^2 \leq s^2$, а отже, заміна загальної дисперсії міжкластерною також може зменшити похибку вибірки. Якщо кластери більш-менш однорідні й варіація кластерних середніх незначна, похибка кластерної вибірки буде меншою порівняно із похибкою простої випадкової чи систематичної вибірки. З іншого боку, кількість дібраних кластерів m значно менша за кількість одиниць спостережень n , що збільшує похибку вибірки. Отже, точність кластерної вибірки залежить як від дисперсії кластерних середніх, так і від кількості дібраних кластерів (серій, гнізд). У кінцевому варіанті похибка кластерної вибірки може перевищувати похибку для простого випадкового чи систематичного добору.

При організації кластерної вибірки, коли кластери різні за обсягом, математичне сподівання вибіркової середньої $E(\bar{x})$ не дорівнює генеральній середній \bar{X} , а перебуває в межах між нею і простою (незваженою) середньою із групових середніх [7]. Розмір систематичної похибки можна визначити за формулою:

$$E(\bar{x}) - \bar{X} \cong \frac{\bar{x} - \bar{X}}{m} \left(\frac{M - m}{M - 1} \right).$$

Незважаючи на це, кластерну вибірку широко застосовують у практиці вибірових обстежень завдяки її відносній дешевизні. Особливо відчутна економія витрат, коли обстеження поширюється на велику територію і кластерами є територіальні одиниці (скорочуються витрати на пересування).

Більшість масштабних вибірових обстежень проводять поетапно із використанням двох і більше способів добору. Особливість багатоступеневої вибірки полягає в тому, що вибіркова сукупність формується поступово, по ступеням, при цьому на кожному ступені своя основа вибірки. Стандартна похибка багатоступеневої вибірки агрегує похибки на всіх ступенях добору і перевищує кожен з них. Так, похибка двоступеневого добору складається із двох стандартних похибок. Припустімо, що за схемою власне випадкової вибірки дібрано n_1 одиниць першого ступеню (гнізд) і з кожної з них — n_2 одиниць другого ступеню. За умови, що гнізда однакового розміру, стандартна похибка середньої визначається за формулою:

$$\mu = \sqrt{\frac{s_B^2}{n_1} + \frac{s_W^2}{n_1 n_2}},$$

де s_B^2 — міжгніздова (міжкластерна) дисперсія;

s_W^2 — середня з гніздових дисперсій;

n_1 — кількість гнізд (кластерів);

n_2 — кількість одиниць, що добираються в кожному кластері;

$n_1 n_2$ — загальна кількість одиниць, дібраних для обстеження.

Для трьох і більше ступенів добору відповідно зростає кількість стандартних похибок, що підлягають агрегації. Розрахунок похибки μ_i ґрунтується на внутрішній варіації відповідної осно-

ви вибірки; формула розрахунку залежить від застосованого способу добору:

$$\mu = \sqrt{\mu_1^2 + \frac{\mu_2^2}{n_1} + \frac{\mu_3^2}{n_1 n_2} + \frac{\mu_4^2}{n_1 n_2 n_3} + \dots},$$

де μ_1, μ_2, μ_3 і т. ін. — стандартні похибки відповідних ступенів добору;

n_1, n_2, n_3 і ін. — обсяги вибірок на відповідних ступенях добору.

Отже, багатоступеневий добір порівняно з іншими видами вибірки менш точний, і кожна нова ступінь добору збільшує сукупну похибку. Водночас цей спосіб має низку переваг в організаційному плані, економічна ефективність його вища, особливо, коли проміжними об'єктами є територіальні одиниці.

Визначення мінімально достатнього обсягу вибірки

Важливим детермінантом точності/надійності результатів спостереження будь-якої вибіркової оцінки є обсяг вибірки. Від обсягу вибірки залежить надійність вартість обстеження. Занадто великий обсяг вибірки потребує зайвих витрат, а надто малий призведе до збільшення похибки репрезентативності. Одним з визначальних етапів планування вибіркового обстеження, передусім формування вибіркової сукупності є розрахунок обсягу вибірки, необхідного для задоволення встановлених вимог щодо надійності оцінок основних показників, бюджету обстеження та навантаження на статистиків, які проводитимуть обстеження.

Теорія вибіркового методу дає змогу науково обґрунтувати мінімально достатній обсяг вибірки, за якого вибіркові оцінки репрезентували б основні властивості генеральної сукупності.

Згідно з формулою граничної похибки $\Delta = t \sqrt{\frac{s^2}{n}}$ обсяг вибіркової сукупності можна визначити як функцію трьох детермінант:

- ступеня однорідності генеральної сукупності s^2 ;

- вимоги до рівня точності вибіркової оцінки, Δ ;
- ймовірності, з якою гарантується результат, t ;

$$n = \frac{t^2 s^2}{\Delta^2}.$$

У тих випадках, коли розрахований обсяг вибіркової сукупності n перевищує 5 % обсягу генеральної сукупності N , його коригують на «безповторність вибірки». Скоригований обсяг вибірки:

$$n' = \frac{n}{1 + \frac{n}{N}}.$$

Практичне використання цих формул ускладнюється через відсутність оцінки варіації. Як правило, використовують оцінки дисперсії s^2 за аналогією, тобто оцінки, отримані в попередніх або аналогічних обстеженнях.

Наприклад, на лісовому масиві у 400 га визначається загальний запас деревини. Пробні ділянки по 0,1 га. За даними попередніх обстежень середнє квадратичне відхилення виходу деревини з 0,1 га становить 3 м³. Скільки пробних ділянок необхідно обстежити, щоб похибка вибірки з імовірністю 0,954 (для якої $t = 2$) не перевищила 1 м³?

Достатній обсяг вибірки пробних ділянок:

$$= \frac{4 \cdot 9}{1^2} = 36.$$

Якщо аналогічні обстеження не проводилися або в генеральній сукупності відбулися істотні зміни, точнішу характеристику варіації дають пробні обстеження невеликої кількості одиниць. За відсутності таких даних оцінку стандартного відхилення як міру варіації можна орієнтовно визначити на підставі варіаційного розмаху, який характеризує діапазон варіації ознаки $R = (x_{max} - x_{min})$ і функціонально пов'язаний зі стандартним відхиленням s :

$$s \approx (x_{max} - x_{min}) : \phi,$$

де ϕ — коефіцієнт, який залежить від обсягу вибірки; значення наведено у табл. 6.5.

Таблиця 6.5

**ЗНАЧЕННЯ КОЕФІЦІЄНТІВ ϕ ДЛЯ ОРІЄНТОВНОЇ
ОЦІНКИ СТАНДАРТНИХ ВІДХИЛЕНЬ**

Обсяг вибірки, n	5	10	20	30	50	≥ 100
Значення ϕ	2,3	3,1	3,7	4,1	4,5	5,0

Для бінарної (альтернативної) ознаки, коли відсутня будь-яка інформація щодо структури сукупності, використовують максимальне значення дисперсії $s^2 = 0,25$.

Залежність обсягу вибірки від заданої точності оцінювання простежимо на прикладі граничної похибки частки Δ_p при ймовірності 0,954 і максимальному рівні варіації (табл. 6.6).

Таблиця 6.6

ЗАЛЕЖНІСТЬ ОБСЯГУ ВИБІРКИ ВІД ТОЧНОСТІ ОЦІНЮВАННЯ

$\Delta_p, \%$	10	9	8	7	6	5	4	3	2	1
Обсяг вибірки	96	118	150	195	266	384	600	1067	2400	9603

Як бачимо, зі збільшенням заданої точності оцінок обсяг вибірки зростає. Так, для забезпечення точності на рівні 10 % необхідно обстежити 96 одиниць, на рівні 9 % — на 22 одиниці більше. Подальше збільшення точності вимагає дедалі більшого обсягу вибіркової сукупності. Абсолютний приріст кількості обстежених на кожен процент приросту точності стрімко зростає (32, 45, 71 і ін.). Отже, при обґрунтуванні точності вибірових даних слід враховувати доцільність додаткових витрат на обстеження.

Обсяг вибірки певною мірою залежить і від масштабу генеральної сукупності. Аби гарантувати стандартну похибку, скажімо, 2,0 %, з генеральної сукупності $N = 1\ 000$ одиниць, в якій частка одиниць з певною ознакою становить $p = 80 \%$, для обстеження необхідно дібрати 286 одиниць, а з сукупності $N = 5\ 000$ за таких самих умов — 370 одиниць. У першому випадку частка вибірки $\frac{n}{N}$ становить 28,6 %, у другому — 7,4 %. Розрахунки виконано за формулою:

$$n = \frac{Nt^2 pq}{N\Delta^2 + t^2 pq}$$

Коли відсутня будь-яка інформація щодо обсягу масштабної генеральної сукупності, визначити обсяг вибірки, який забезпечить належний рівень точності вибіркових оцінок, можна за формулою:

$$\lim_{N \rightarrow \infty} n = \lim_{N \rightarrow \infty} n \frac{pq}{\mu_p^2 + \frac{pq}{N}} = \frac{pq}{\mu_p^2}.$$

Для невеликих за обсягом сукупностей застосування цієї формули неефективне.

Точність оцінювання як основа визначення обсягу вибірки встановлюється споживачами результатів обстеження у вигляді абсолютної або відносної граничної похибки (для прийнятної ймовірності). Спеціалісти, проектуючи вибіркове обстеження, зазвичай контролюють відносну граничну похибку V_{Δ} . У такому разі і варіація ознаки вимірюється відносною величиною. Для ознак метричної шкали це коефіцієнт варіації V_x , який розраховується відношенням стандартного відхилення до середньої величини:

$$n = \frac{t^2 V_x^2}{V_{\Delta}^2},$$

де V_x — коефіцієнт варіації; V_{Δ} — гранична відносна похибка вибірки.

Обидві характеристики виражаються у процентах.

Наприклад, проектується вибіркове обстеження підприємств IT-галузі регіону ($N = 125$) з метою визначення середнього розміру чистого прибутку. За даними аналогічного обстеженнями, чистий прибуток становив 42 млн грн, квадратичний коефіцієнт варіації $V_x = 20\%$. Мінімально необхідний обсяг вибірки, при якому з імовірністю 0,954 гарантується відносна похибка вибірки розміром не більш як 8 %, становить:

$$n = \frac{2^2 \cdot 20^2}{8^2} = 25.$$

Скоригований на скінченність сукупності обсяг вибірки дещо менший:

$$n' = \frac{25}{1 + \frac{25}{125}} = \frac{25}{1,2} \approx 21.$$

Аналогічно можна визначити достатній обсяг вибірки за заданою відносною похибкою частки:

$$n = \frac{t^2 q}{V_{\Delta}^2 p}.$$

Очевидно, чим більша частка p , тим менший обсяг вибірки забезпечить необхідну точність результатів обстеження, і навпаки, для малих значень p обсяг вибірки збільшується.

У табл. 6.7 наведено обсяги вибірки, які забезпечують точність результатів обстеження малопоширених явищ з відносною стандартною похибкою, меншою за 10 %.

Таблиця 6.7

ДОСТАТНІЙ ОБСЯГ ВИБІРКИ ДЛЯ ВИВЧЕННЯ МАЛОПОШИРЕНИХ ЯВИЩ

p	q / p	n при $V_{\mu} \leq 10\%$
0,20	4,0	400
0,15	5,7	570
0,12	7,3	730
0,10	9,0	900
0,09	10,1	1010
0,08	11,5	1150

Мінімально достатній обсяг вибірки зі складним дизайном $n(D)$ визначається з урахуванням відповідного дизайн-ефекту:

$$n(D) = n \text{ deff} \quad \text{або} \quad n = \text{deff} \cdot s^2 \left\| \frac{100}{V_{\mu} \times \bar{y}} \right\|^2,$$

де s^2 — дисперсія, яка характеризує варіацію значень показника у вибірці, сформованій за принципом простого випадкового добору;

V_{μ} — відносна похибка, визначена у вимогах до формування вибірки;

\bar{y} — оцінка середнього значення показника.

За необхідності отримання надійних оцінок для часток та сумарних значень обсяги вибірок зі складним дизайном розрахо-

вуються за аналогічними, дещо модифікованими формулами — оцінка середнього значення замінюється відповідно на значення частки або сумарного значення показника:

$$\begin{array}{ll} \text{для частки } p: & \text{для сумарного показника } Y: \\ n = deff \frac{(1-p)}{p} \cdot \left(\frac{100}{v_{\mu}} \right)^2, & n = deff \cdot N^2 \cdot s^2 \left(\frac{100}{v_{\mu} \cdot Y} \right)^2. \end{array}$$

Якщо $deff < 1$, то бажану точність результатів обстеження забезпечить вибірка меншого обсягу — $(n(D) < n)$ і навпаки.

На жаль, значення $deff$, як і дисперсія, невідомі до проведення обстеження, їх можна оцінити лише за даними вибірки. Тому на практиці при розрахунку обсягу вибірки використовують непрямі оцінки значення $deff$, визначені за результатами аналогічних чи пробних обстежень.

Оскільки базові обстеження проводяться на постійній основі (щомісячно, щоквартально, щорічно) і дизайн вибірки не зазнає суттєвих змін, фактичний обсяг вибірки такого обстеження може бути скоригований на результати попередніх обстежень або вимог до проведення поточного обстеження. Наприклад, при визначенні загального обсягу вибірки на наступний рік з урахуванням вимог щодо рівня надійності вибіркової оцінки:

$$n_{st} = \left(\sum_j n_j \frac{v_j^2}{v_{st}^2} \right),$$

де n_{st} — теоретичний загальний обсяг вибірки з урахуванням вимог;

n_j — фактичний обсяг вибірки за результатами попереднього циклу обстеження;

$V_{\mu,j}$ — рівень надійності оцінок за попереднім циклом обстеження;

$V_{\mu,st}$ — рівень надійності оцінок, передбачений вимогами, скажімо, для j -го регіону на рівні 10 %.

Під час проведення вибірових обстежень певна частина респондентів з різних причин відмовляється брати участь в обстеженні. Відмови на рівні 10 % можуть привести [Гладин, 2008:

с. 63] як до зменшення надійності, так і до зміщення оцінок показників. Задля зменшення впливу відмов респондентів від участі в обстеженні на точність і надійність вибірових оцінок здійснюють коригування визначеного обсягу вибірки на очікуваний коефіцієнт відповідей:

$$n_{rr} = \frac{n}{k_{rr}},$$

де n_{rr} — обсяг вибірки з урахуванням відмов;

k_{rr} — коефіцієнт участі респондентів в обстеженні, визначений експертами на підставі досвіду проведення аналогічних обстежень, за результатами пілотних обстежень або за експертними оцінками (коефіцієнт відмов дорівнює $(1 - k_{rr})$).

У разі, коли необхідно отримати надійні оцінки показників для окремих сегментів об'єкта дослідження, наприклад, для регіонів, обсяг вибірки визначається відповідно для кожного регіону. Сума обсягів вибірок за всіма регіонами є загальним підсумковим обсягом вибірки для країни загалом.

У більшості вибірових обстежень одночасно вивчається не одна, а кілька ознак. Якщо бажаний рівень точності визначати для кожної ознаки окремо, то результатом розрахунків стане низка значень обсягів вибірки. З урахуванням бюджету обстеження вибирають або максимальний обсяг n (і тоді решта ознак оцінюється «надто точно»), або за основу беруть точність оцінки головної ознаки.

У практиці вибірових обстежень рішення про обсяг вибірки є компромісом між теоретичними припущенням щодо точності результатів і можливостями їх практичної реалізації. Насамперед ідеться про баланс між обсягом вибірки, вимогами щодо надійності оцінок і витратами на обстеження. Розрахований для заданої надійності оцінок обсяг вибіркової сукупності порівнюється з фінансовими можливостями замовника. Визначається необхідна кількість статистиків, які мають проводити польовий етап обстеження, інших працівників, які забезпечуватимуть розроблення й аналіз результатів, розраховується також необхідний обсяг фінансових витрат. Коли бюджет дослідження обмежений і забезпечити дотримання певних умов формування вибірки важко, слід здійснити коригування вимог, скажімо, в плані зменшення надійності вибірових оцінок. Після цього розраховується новий

обсяг вибіркової сукупності і процедура порівняння вимог з можливостями здійснюється повторно. Таким чином, процедура визначення обсягу вибірки може бути ітераційною і завершиться після узгодження всіх вимог до вибірки з фінансовими й інституціональними можливостями замовника.

Вправи та запитання для самоконтролю

1. Поясніть концептуальні засади оцінювання точності вибірових оцінок.

2. Як визначити розмір похибки вибірки? Чим відрізняється гранична похибка вибірки від стандартної (середньої)?

3. Що таке незміщеність оцінки параметра вибірки?

4. Що таке ефективність оцінки параметра вибірки?

5. У чому особливість малої вибірки?

6. Як зміниться похибка вибірки, якщо обсяг вибіркової сукупності збільшиться в 2,25 рази?

7. Як позначиться на стандартній похибці вибірки збільшення дисперсії в 1,7 рази?

8. Як зміниться стандартна похибка вибірки, якщо замість простого випадкового добору виконати 19%-й систематичний?

9. Поясніть суть і функції граничної похибки вибірки.

10. Як зміниться стандартна похибка вибірки, якщо замість механічного добору виконати стратифікований з міжгруповою дисперсією, що становить 36 % від загальної?

11. Поясніть особливості розрахунку похибки кластерної вибірки.

12. Як визначити похибку багатоступеневої вибірки?

13. З метою визначення загального обсягу деревини в сосновому лісі обстежено 100 дерев. Середній діаметр дерев становив 35 см при дисперсії 144. З імовірністю 0,954 визначте граничну похибку для середнього діаметра дерев.

14. За даними технічного аналізу 100 проб руди вміст заліза становить у середньому 60 % при стандартній похибці вибірки 4,5 %. Чи є підстави стверджувати з імовірністю 0,954, що для цього родовища вміст заліза в руді не менше 51 %?

15. За результатами контрольної перевірки податковими службами 400 бізнесових структур 140 з числа перевірених у податкових деклараціях за минулий рік вказали не всі доходи, які підлягають оподаткуванню. Визначте частку бізнесових структур, які приховують частину доходів від сплати податків, та довірчі межі частки з імовірністю 0,954.

16. За даними опитування 100 підприємців (19 %-а вибірка) 20 з них оцінюють економіко-правові умови їхньої діяльності як несприят-

ливі. З імовірністю 0,954 визначте похибку вибірки для частки респондентів, не задоволених умовами діяльності.

17. За даними вибіркового обстеження втрат зерна озимої пшениці через несвоєчасне збирання врожаю середній рівень втрат становив 3 ц/га, коефіцієнт варіації — 30 %. Обстежено 100 ділянок. З імовірністю 0,954 визначте відносну похибку вибірки для середнього рівня втрат врожаю.

18. За даними обстеження 100 домогосподарств середньодушкове споживання поживних речовин за добу становить 3200 ккал за стандартної похибки вибірки 25,6 ккал, а білків — 90 г за стандартної похибки 1,35 г. Для якого показника точність вибірки вища?

19. На зерноприймальному пункті для встановлення режиму роботи зерносушарки перевіряється вологість зерна взяттям вибірково дібраних проб з автомашин із зерном, що надійшло від двох виробників. За результатами перевірки виявилось, що середня вологість зерна виробника А становить 6 % при дисперсії 25, виробника Б — 10 % при дисперсії 36. Для якого виробника похибка вибірки вологості зерна менша?

20. Під час аудиторської перевірки фінансової діяльності транспортної компанії з метою визначення швидкості розрахунків із кредиторами було випадково дібрано 100 платіжних документів. Згідно з перевіреними документами середній термін перерахунку й отримання коштів 27 днів з дисперсією 36. Кожен десятий платіжний документ був з порушенням терміну розрахунків.

Визначте стандартні похибки для середнього терміну проходження платежів упродовж року і частки платежів з порушенням термінів розрахунку. Порівняйте точність оцінок цих показників за допомогою відносних похибок вибірки.

21. За результатами обстеження 100 проб вугілля встановлено, що середня зольність вугілля в родовищі 16 % при $\sigma = 5$ %. При цьому в 10 % проб зольність вугілля перевищила 20 %. З імовірністю 0,954 визначте довірчі межі для середньої зольності вугілля і для частки вугілля, що має зольність понад 20 %. Порівняйте точність оцінок цих показників за допомогою відносних похибок вибірки.

22. За даними вибіркового обстеження посівів озимої пшениці визначена частка зимової загибелі їх:

Сорт пшениці	Обстежена площа, га	Частка зимової загибелі посівів, %
А	6	10
Б	4	15

Визначте середню частку зимової загибелі посівів і граничну похибку вибірки для частки з імовірністю 0,954.

23. Дирекція мережі супермаркетів провела вибіркове опитування працівників двох регіонів щодо підтримки ними нової маркетингової політики:

Регіон	Опитано	Підтримують, %
А	24	70
Б	16	80

Визначте середній рівень підтримки нової маркетингової політики і граничну похибку вибірки цього показника з імовірністю 0,954.

24. Як порівняти точність та надійність результатів вибіркового обстеження за різними планами вибірки?

25. Визначте дизайн-ефект систематичної вибірки, якщо планується обстеження 7,5% генеральної сукупності.

26. Поясніть зв'язок між обсягом вибірки і точністю, з якою гарантуються результати обстеження.

27. Хімічний аналіз 25 партій молока дав такі результати:

а) середній показник кислотності (у градусах Тернера) — 20⁰ при дисперсії 3,24;

б) частка партій молока, що відповідають стандарту кислотності (не більше 21⁰) — 80 %.

Визначте похибки вибірки для середньої і частки з імовірністю 0,954.

Скільки партій молока необхідно перевірити, щоб похибки вибірки для середньої і частки з тією самою ймовірністю зменшити вдвічі?

28. Необхідно дослідити середню вагу форелі, вирощуваної у штучному водоймі. Скільки рибин необхідно відловити для обстеження, щоби з ймовірністю 0,954 помилитися у вазі форелі не більше, ніж на 20 г, якщо відомо, що за аналогічних умов вирощування форель має середню вагу 500 г з дисперсією 2550 г?

29. За даними попередніх вибірових обстежень частка нестандартної плодоовочевої продукції становила 0,10. Скільки треба провести перевірок, щоб гранична похибка вибірки частки з імовірністю 0,954 не перевищила 4 %?

30. Планується вибіркове обстеження мешканців району з метою вивчення заборгованості за житлово-комунальні послуги. За схемою повторного добору визначено обсяг вибірки — 243 домогосподарства. Скоригуйте обсяг вибірки на «скінченність сукупності» за умови, що частка вибірки становить 8 %.

31. Вивчається міра пошкоджень дерев парку короїдами. Скільки дерев необхідно обстежити, щоб з імовірністю 95 % помилитись не бі-

льше ніж на 5 %? Відомо, що в сусідньому парку частка пошкоджених дерев становить 3 %.

32. Скільки необхідно перевірити проб вугілля, що надійшло на електростанцію, щоб похибка вибірки для середнього рівня зольності з імовірністю 0,954 не перевищила 5 %? За даними попередніх перевірок коефіцієнт варіації зольності вугілля становить 20 %.

33. Проектується обстеження посівів соняшника з метою визначення втрат насіння через несвочасне збирання врожаю. Скільки необхідно обстежити пробних ділянок, щоб з імовірністю 0,954 відносна похибка середніх втрат з 1 га не перевищила 10 %? За даними минулорічних обстежень квадратичний коефіцієнт варіації втрат насіння становив 20 %.

34. Скільки треба опитати респондентів, оцінюючи якість роботи органів місцевого самоврядування (задовольняє / не задовольняє), щоб гранична похибка вибірки часток з імовірністю 0,954 при цьому не перевищила 5 %?

35. Організується моментне спостереження за ефективністю використання 25 одиниць виробничого устаткування. Скільки спостережень необхідно провести, щоб похибка вибірки для частки устаткування, що не працює, з імовірністю 0,954 не перевищила 2 %? За умови, що спостереження триватиме 5 днів, скільки спостережень щоденно слід провести за кожною одиницею устаткування?

36. Проектується обстеження трудової активності сільського населення в регіоні. За аналогічними обстеженнями в інших регіонах частка зайнятих чоловіків працездатного віку становить 70 %, жінок — 50 %. Якою має бути чисельність вибірки, щоб похибка вибірки для частки зайнятого населення з імовірністю 0,954 не перевищила 5 %?

Рекомендована література

1. *Гладун О.В., Вацаєва Н.А.* Дизайн вибірки та проблеми його оптимізації // Проблеми теорії та методології бухгалтерського обліку, контролю і аналізу : Зб. наук. пр. — Житомир: ЖДТУ, 2007, вип. 2(8). — С. 31–39.

2. *Єріна А.М.* Організація вибірових обстежень : Навчальний посібник — К : КНЕУ, 2003. — 127 с.

3. *Кокрен У.* Методы выборочного обследования. М. : Статистика, 1976. — 440 с.

4. Статистика підприємств : Навчальний посібник / За ред. С.О. Матковського. — К. : Алерта, 2017. — 560 с.

5. *Черняк О.І.* Техніка вибірових досліджень. — К : МІВВЦ, 2001. — 248 с.

РОЗДІЛ 7

ФОРМУВАННЯ СТРАТИФІКОВАНОЇ ВИБІРКИ

Методологічні принципи стратифікації
Способи розміщення (алокації) стратифікованої вибірки
Стратифікація у практиці державних вибіркових обстежень

Після вивчення цього розділу студент має

- *знати*: логіку і методологічні принципи стратифікації, застосування яких дозволяє підвищити надійність результатів вибірки; способи розміщення вибіркової сукупності за стратами і областями дослідження;
- *уміти*: відповідно до мети дослідження обирати параметри стратифікації і оцінювати дизайн-ефекти стратифікованої вибірки; застосовувати ефективні способи розміщення вибірки за стратами і областями дослідження, враховуючи кошторис обстеження; оцінювати точність результатів обстеження.

Методологічні принципи стратифікації

Широке впровадження вибіркових обстежень в економіку та соціологію, де об'єкти переважно своїй неоднорідні і поділяються на певні сегменти, сприяло розробленню таких схем формування вибірки, за якими кожен сегмент генеральної сукупності був би представлений власною вибіркою.

Як зазначалося у 5.3, визначальним методологічним принципом організації вибіркових обстежень неоднорідних сукупностей є *стратифікація*, сутність якої полягає в поділі генеральної сукупності на певні групи (страти, домени) і доборі одиниць у вибірку сукупність з кожної страти. Ознаки, покладені в основу поділу сукупності на групи, називають параметрами стратифікації. Так, у вибіркового обстеженні малих підприємств основними параметрами стратифікації є вид економічної діяльності та розмір підприємств, у сільському господарстві страти можуть бути представлені природно-економічними зонами, в соціологічних опитуваннях — типами поселень та соціально-демографічними групами населення.

Головне призначення стратифікації — подолати умови неоднорідності сукупності. Чим одноріднішим є склад утворених страт, тим меншою є внутрішньокласова варіація і вищим рівень репрезентативності вибірових оцінок. Стосовно кількості страт, які б виявили відмінності середніх, універсального правила немає. У кожному обстеженні це питання вирішується залежно від особливостей структури сукупності і взаємозв'язків цільових характеристик.

Стратифікацію здійснюють на різних етапах формування вибірки, її об'єктами можуть бути і безпосередні (кінцеві) одиниці спостереження, і проміжні вибірові одиниці (міста, фірми, школи тощо). Найчастіше стратифікації піддаються проміжні вибірові одиниці, скажімо, в соціально-демографічних обстеженнях міста поділяються на великі й малі.

Практично процедура стратифікації спирається на групування одиниць генеральної сукупності за однією ознакою чи за комбінацією ознак. Тому далі всі параметри, що характеризують окремі страти, будемо називати груповими.

Стратифікована вибірка добувається сепаратно із кожної ідентифікованої чи утвореної j -ої групи ($j = 1, 2, \dots, m$) в обсязі n_j одиниць. Добір організується або пропорційно чисельності груп, або пропорційно варіації ознаки всередині груп. Сума дібраних за групами одиниць формує загальний обсяг вибірки $n = \sum_{j=1}^m n_j$.

За даними вибірок визначають групові середні $\bar{x}_j = \frac{1}{n_j} \sum_{h=1}^{n_j} x_{jh}$ і на їх основі — загальну середню для сукупності загалом як арифметичну зважену із групових середніх:

$$\bar{x} = \frac{\sum_{j=1}^m \bar{x}_j N_j}{N} = \sum_{j=1}^m \omega_j \bar{x}_j,$$

де $\frac{N_j}{N} = \omega_j$ — вага j -ої групи в генеральній сукупності.

Аналогічно співвідносяться квадрати стандартних похибок загальної і групових середніх:

$$\mu_{\bar{x}}^2 = \sum_{j=1}^m \omega_j^2 \mu_j^2 = \sum_{j=1}^m \omega_j^2 \frac{s_j^2}{n_j} (1 - f_j),$$

де $\mu_j^2 = \frac{s_j^2}{n_j}(1 - f_j)$ — похибка групової середньої;

s_j^2 — оцінка групової дисперсії (для частки $s_j^2 = p_j q_j$);

$f_j = \frac{n_j}{N_j}$ — частка добору всередині j -ої групи.

За умови, що частки добору в групах фіксовані, тобто $f_j = f$ або $\frac{n_j}{N_j} = \frac{n}{N}$, стратифікована вибірка буде *самозваженою*, а точність

її оцінюється за формулою

$$\mu_x^2 = \frac{1-f}{n} \sum_j \omega_j s_j^2.$$

Якщо частки добору в групах незначні ($f_j < 5\%$), коригувальний коефіцієнт $(1 - f_j)$ у формулу похибки вибірки не вводиться. Як зазначено в 6.3, середню з групових дисперсій можна визначити, скоригувавши загальну дисперсію на щільність кореляційного зв'язку, тобто $s_w^2 = s^2(1 - r^2)$. Звідси:

$$\mu_{start}^2 = \frac{s^2(1 - r^2)}{n},$$

де r — коефіцієнт кореляції.

Як бачимо, розшарування сукупності зменшує похибку вибірки на частку $(1 - r^2)$. Чим щільніший зв'язок між ознаками, тим помітніше зменшення похибки. При $r^2 = 0,50$ похибка вибірки зменшується вдвічі, при $r^2 = 0,66$ — втричі.

Отже, якщо для простої випадкової вибірки єдиний спосіб вплинути на точність вибірових оцінок — збільшити чи зменшити обсяг вибірки, то в стратифікованій вибірці визначальним детермінантом є зменшення внутрішньогрупової варіації. Дизайн-ефект стратифікованої вибірки порівняно з простою випадковою при обстеженні одного й того самого об'єкта менший за одиницю:

$$deff_{strat} = \frac{\mu_{strat}^2}{\mu^2} \approx \frac{\sum_{j=1}^m s_j^2 n_j}{s^2 n} < 1.$$

Із наведених формул можна бачити, що точність і надійність оцінок стратифікованої вибірки залежить від того, наскільки виокремленні групи однорідні, а випадково дібрані сукупності n_j репрезентують відповідні групи. Досягти значного ефекту стратифікації можна за умови формування груп, які сприяють зменшенню внутрішньогрупової варіації і посилюють міжгрупові відмінності. Першою умовою є вдалий вибір ознаки стратифікації, тобто ознаки, що визначає належність одиниці сукупності до певної групи. Як доведено у 6.3, внутрішньогрупова варіація буде тим меншою, чим щільнішим є зв'язок між параметром стратифікації і цільовою ознакою. Найбільший ефект дає стратифікація за розміром об'єктів дослідження (компаній — за розміром активів, агроферм — за площею сільськогосподарських угідь, магазинів — за обсягом товарообігу тощо).

При оцінюванні часток, якщо вони не надто змінюються від групи до групи, ефект стратифікації виявляється незначним. Річ у тім, що дисперсія частки $s_j^2 = pq$ мало чутлива до міжгрупових розбіжностей: групові частки відхиляються від загальної в обидва боки і дисперсія зменшується непомітно. У табл. 7.1 наведено відносну точність стратифікованої вибірки порівняно із простою випадковою у трьох групах однакового розміру [Кокрен, 1976: с. 124]. Розглядаються чотири варіанти співвідношення часток за групами: за першим варіантом частки змінюються в межах 0,4–0,6, за останнім — у межах 0,1–0,9. Як видно з даних таблиці, помітний вираш у точності виявляють лише ті групи, де варіація часток є надто високою і практично малоімовірною. За реальної варіації часток (1-а група) ефект стратифікації становить лише 3 %.

Таблиця 7.1

ВІДНОСНА ТОЧНІСТЬ СТРАТИФІКОВАНОЇ І ПРОСТОЇ ВИПАДКОВОЇ ВИБІРКИ ПРИ ОЦІНЮВАННІ ЧАСТОК

Значення часток у стратах	Дисперсія часток		Відносна точність, у %, $100 s^2 / s_{strat}^2$
	Проста вибірка $s^2 = pq$	Стратифікована вибірка $s_{strat}^2 = \frac{1}{3} \sum p_j q_j$	
0,4; 0,5; 0,6	0,25	0,2433	103
0,3; 0,5; 0,7	0,25	0,2233	122
0,2; 0,5; 0,8	0,25	0,1900	132
0,1; 0,5; 0,9	0,25	0,1433	174

Не виявляє очікуваного ефекту також стратифікація за територіальними чи організаційними ознаками, якщо межі груп установлені адміністративно (райони області, квартали міста тощо) і рівень їхньої однорідності не піддається регулюванню. Варіація всередині таких «районів» може виявитися більшою, ніж «між-районна», а отже, стратифікована вибірка втрачає свої переваги щодо простої випадкової.

Друга умова ефективної стратифікації пов'язана з розміщенням вибірки за стратами. Точність і надійність вибіркових оцінок залежать не лише від того, за якою ознакою здійснена страифікація, а й від того, в який спосіб розподілено вибіркову сукупність за виокремленими стратами. Вдалий розподіл сукупності на однорідні страти дає змогу відобразити специфіку досліджуваного об'єкта й уникнути зміщення вибіркових оцінок.

Отже, процедура формування стратифікованої вибірки реалізує фундаментальний принцип рандомізації і забезпечує обґрунтованість та «законність» висновків про генеральну сукупність, зроблених на підставі обстеження невеликих за обсягом сукупностей.

Способи розміщення (алокції) стратифікованої вибірки

Важливим етапом організації стратифікованої вибірки є розміщення вибіркової сукупності за стратами, тобто алокація обсягу вибірки $n = \sum n_j$. У практиці вибіркових обстежень застосовують різні способи розміщення стратифікованої вибірки. Найпростіший з них — з кожної страти добрати однакову кількість одиниць:

$$n_j = \frac{n}{m} \cdot$$

Проте застосування такого простого способу обмежене. В разі, коли обсяг генеральної сукупності великий, а частоти окремих страт помітно різняться, може виникнути ситуація, коли $n_j > N_j$. Аби задовольнити умову $n_j \leq N_j$, на практиці для розміщення вибірки за стратами застосовують такі способи, як пропорційний та оптимальний.

Пропорційний розподіл передбачає однакове представництво для всіх сегментів сукупності, тобто частки добору в стратах і загальною по сукупності однакові $\frac{n_j}{N_j} = \frac{n}{N}$, а обсяг вибірки n_j залежить від обсягу відповідної групи в генеральній сукупності:

$$n_j = \frac{N_j}{N} n.$$

Наприклад, здійснюється вибіркове опитування кожного п'ятого безробітного, зареєстрованого в службах зайнятості міста. Всього зареєстровано 200 осіб, з них 30 % мають вищу освіту (список *A*), 25 % — середню спеціальну (список *B*) і 45 % — середню і неповну середню (список *B*). Вибіркова сукупність має бути $200 : 5 = 40$ осіб. Очевидно, зі списку *A* буде відібрано $(40 \times 0,3) = 12$ осіб, зі списку *B* — $(40 \times 0,25) = 10$, зі списку *B* — $(40 \times 0,45) = 18$.

Якщо структура генеральної сукупності відома, то практична реалізація пропорційного добору нескладна. На жаль, такий добір не завжди може забезпечити високу репрезентативність вибірко-вих даних. Розподіл первинної сукупності часто, особливо в соціально-економічних дослідженнях, асиметричний — із правосторонньою асиметрією. У такому разі групи з високими значеннями показників малочисельні і за пропорційного добору довелося б добирати з цих груп дуже мало одиниць. Водночас у малочисельних групах варіація значно вища порівняно з іншими групами. Очевидно, що при розміщенні вибірки за групами слід враховувати ступінь варіації, тобто частка вибірки має бути більшою там, де більша варіація. Це так зване *оптимальне* розміщення стратифікованої вибірки (розміщення Неймана). Оптимальним у плані мінімізації похибки є обсяг групи, пропорційний до середнього квадратичного відхилення в групі — s_j :

$$n_j = \frac{N_j s_j}{\sum_1^m N_j s_j} n,$$

де n — обсягу вибірки, m — кількість груп.

Порівняймо точність пропорційного і оптимального розміщень 20 %-вибірки обсягом $n = 50$. У табл. 7.2 наведено розподіл генеральної сукупності за групами ($m = 4$).

Пропорційне розміщення вибірки. Добір одиниць із j -ої групи здійснюють пропорційно груповим частотам первинної сукупності, тобто:

$$n_1 = 50 \times 0,72 = 36; n_2 = 50 \times 0,17 = 8 \text{ і т. д.}$$

Значення групових середніх \bar{x}_j свідчать про наявність зв'язку між ознакою стратифікації і ознакою, параметри якої оцінюються: групові середні відрізняються від загальної ($\bar{x} = 3,1$) і систематично від групи до групи зростають. Водночас нерівномірною виявилася внутрішньогрупова варіація — групові дисперсії s_j^2 істотно різняться.

Таблиця 7.2

ПРОПОРЦІЙНЕ РОЗМІЩЕННЯ СТРАТИФІКОВАНОЇ ВИБІРКИ

Показник	Група				По сукупності загалом
	1	2	3	4	
N_j	180	42	20	8	250
$N_j / \sum N_j$	0,72	0,17	0,08	0,03	1,0
Середня \bar{x}_j	2,7	3,5	4,9	5,6	3,1
Дисперсія s_j^2	0,348	0,922	1,742	3,204	0,602
Пропорційне розміщення — n_j	36	8	4	2	50

Визначимо середню з групових і міжгрупову дисперсії ознаки x . Середня з групових дисперсії становить 0,602:

$$s_w^2 = \frac{\sum_j s_j^2 n_j}{\sum_j n_j} = \frac{0,348 \cdot 180 + 0,922 \cdot 42 + 1,742 \cdot 20 + 3,204 \cdot 8}{250} = 0,602,$$

а міжгрупова дисперсія — 0,601:

$$s_B^2 = \frac{\sum_1^m (\bar{x}_j - \bar{x})^2 n_j}{\sum_1^m n_j} = \frac{(2,7-3,1)^2 \cdot 180 + (3,5-3,1)^2 \cdot 42 + (4,9-3,1)^2 \cdot 20 + (5,6-3,1)^2 \cdot 8}{250} = 0,601.$$

За правилом декомпозиції варіації загальна дисперсія дорівнює:

$$0,601 + 0,602 = 1,203.$$

Порівняння дисперсії показує, наскільки ефективною виявилася стратифікація: внутрішньогрупова дисперсія, яка лежить в основі розрахунку похибки стратифікованої вибірки, майже удвічі менша за загальну дисперсію, яка є основою розрахунку похибки простої випадкової вибірки.

Стандартна похибка стратифікованої вибірки для середньої становить:

$$\mu_{\bar{x}} = \sqrt{\frac{0,602}{50} (1 - 0,8)} = 0,0981.$$

При простому випадковому доборі стандартна похибка була б

$$\mu_{\bar{x}} = \sqrt{\frac{1,203}{50} (1 - 0,8)} = 0,1387.$$

Як свідчить *design-effect*, стратифікація вибірки підвищила точність результатів спостереження порівняно з простою вибіркою майже удвічі:

$$deff_{strat} = 0,0981^2 : 0,1387^2 = 0,00962 : 0,01924 \approx 0,5.$$

Оптимальне розміщення. Оскільки варіація за групами істотно різниться, визначимо обсяги вибірки з кожної групи пропорційно варіації. За даними табл. 7.3 сумарна варіація становить

$$\sum_1^4 N_j s_j = 187,2.$$

Добір одиниць із j -ої групи здійснюють пропорційно груповій варіації:

$$n_1 = (106,2 : 187,2) \times 50 = 28;$$

$$n_2 = (40,3 : 187,2) \times 50 = 11 \text{ і т. д.}$$

Оптимальне розміщення вибірки подано в останньому рядку табл. 7.3.

Таблиця 7.3

ОПТИМАЛЬНЕ РОЗМІЩЕННЯ СТРАТИФІКОВАНОЇ ВИБІРКИ

Показник	Група				По сукупності загалом
	1	2	3	4	
N_j	180	42	20	8	250
Середня в j -ій групі, \bar{x}_j	2,7	3,5	4,9	5,6	3,1
Стандартне відхилення в j -ій групі, s_j	0,59	0,96	1,32	1,79	X
$N_j s_j$	106,2	40,3	26,4	14,3	187,2
$N_j s_j / \sum N_j s_j$	0,57	0,22	0,14	0,07	1,0
Оптимальне розміщення, n_j	28	11	7	4	50

Відхилення оптимального розміщення від пропорційного унаочнено на рис. 7.1. Однорідні групи (з меншою варіацією) при оптимальному розподілі подаються меншою кількістю вибірових одиниць, неоднорідні (з більшою варіацією) — більшою.

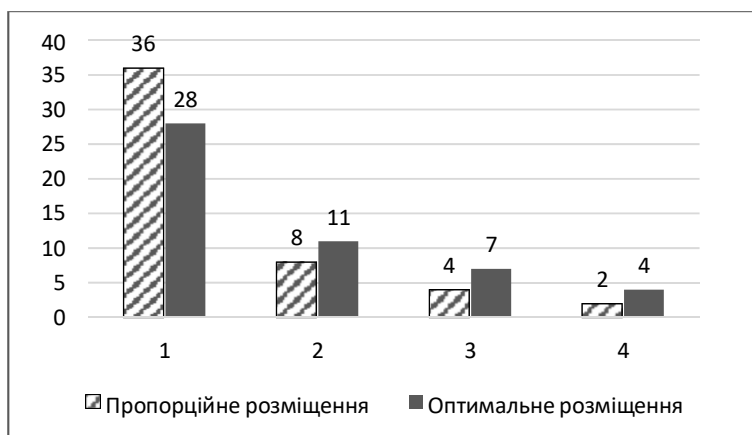


Рис. 7.1. Пропорційне і оптимальне розміщення стратифікованої вибірки

Оптимальне розміщення дає значний виграш у точності результатів обстеження у двох випадках:

- коли сукупність об'єднує великі і малі об'єкти, а розшарування здійснюється за їхнім розміром (великі та малі міста), тоді дисперсія для великих об'єктів значно більша і пропорційне розміщення неефективне;

- коли витрати на обстеження в одних сегментах сукупності набагато більші, ніж в інших.

Практична реалізація оптимальної стратифікованої вибірки стикається з певними труднощами, передусім через відсутність даних про варіацію ознаки у виокремлених групах. Через похибки при оцінюванні дисперсії відносна надійність оптимальної вибірки може виявитися дещо перебільшеною.

Отже, під час проектування вибірки віддавати перевагу оптимальному розміщенню доцільно тоді, коли є необхідна інформація про варіацію ознаки стратифікації і коли дисперсії цієї ознаки в окремих стратах s_j^2 помітно різняться. Якщо ж групові дисперсії мало відрізняються одна від одної і за значенням близькі до загальної дисперсії ($s_j^2 \approx s^2$), то оптимальне розміщення дає таку саму точність, як і пропорційне. Передусім це стосується оцінювання часток. Відносна точність пропорційного розміщення порівняно з оптимальним при оцінюванні часток визначається відношенням:

$$\frac{\mu_{opt}^2}{\mu_{prop}^2} = \frac{(\sum w_j \sqrt{p_j q_j})^2}{\sum w_j p_j q_j},$$

де w_j — вага j -ої страти.

Відносна точність пропорційного добору щодо оптимального для двох груп однакового розміру в інтервалі $[p \div (1 - p)]$ можна простежити за даними табл. 7.4 [Кокрен, 1976: с. 125]. Як бачимо, навіть при $p = 0,1$ або $0,9$ відносна точність становить лише 94 %. Очевидно, що простота пропорційного розміщення і його властивість самозважування може компенсувати таку невелику втрату точності.

Таблиця 7.4

**ВІДНОСНА ТОЧНІСТЬ ПРОПОРЦІЙНОГО РОЗМІЩЕННЯ
ПОРІВНЯНО З ОПТИМАЛЬНИМ ПРИ ОЦІНЮВАННІ ЧАСТКИ**

P	0,4 або 0,6	0,3 або 0,7	0,2 або 0,8	0,1 або 0,9	0,05 або 0,95
Відносна точність, %	100,0	99,8	98,8	94,1	86,6

Стратегія розміщення стратифікованої вибірки певною мірою визначається витратами на обстеження C . Нехай одиничні витрати в j -й страті становлять c_j , тоді вартість обстеження дібраних з цієї страти одиниць n_j буде $c_j n_j$, а загальна вартість обстеження:

$$C = C_0 + \sum_1^m c_j n_j,$$

де C_0 — витрати, що не залежать від обсягу вибірки.

Коли бюджет обстеження обмежений, мінімум витрат ($C - C_0$) забезпечить той варіант розміщення вибірки, в якому:

$$n_j = n \frac{N_j s_j / \sqrt{c_j}}{\sum_1^m (N_j s_j / \sqrt{c_j})},$$

тобто коли обсяги страт n_j пропорційні множнику $\frac{N_j s_j}{\sqrt{c_j}}$. Відтак,

обсяг стратифікованої вибірки необхідно розподіляти на користь великих, більш варіабельних і дешевших за витратами страт. За умови фіксованих витрат ($C - C_0$) і оптимального розміщення вибірки за стратами загальний обсяг вибірки n становить:

$$n = \frac{(C - C_0) \sum_1^m (N_j s_j / \sqrt{c_j})}{\sum_1^m (N_j s_j / \sqrt{c_j})}.$$

Завдання полягає в тому, щоб підібрати такі обсяги вибірки n_j , які б у межах означеної вартості ($C - C_0$) мінімізували стандартну похибку вибірки μ_x або мінімізували витрати за певного рівня похибки.

Частка j -ої страти у вибірці залежить від обсягу страти в генеральній сукупності N_j , варіації ознаки s_j і витрат на обстеження вибіркової одиниці c_j :

$$\frac{n_j}{n} = \frac{N_j s_j / \sqrt{c_j}}{\sum_1^m (N_j s_j / \sqrt{c_j})}.$$

Щоб застосувати наведені формули, треба знати не лише групові дисперсії, а й одиничні витрати на обстеження c_j в усіх стратах. Очевидно, більша кількість одиниць n_j припаде на ті страти, для яких одиничні витрати на обстеження менші, і навпаки. Якщо одиничні витрати c_j однакові в усіх стратах, отримаємо оптимальне розміщення.

У разі стратифікації сукупності за двома і більше ознаками найкраще розміщення за однією ознакою не буде найкращим за іншою. У таких випадках приймають *компромісне* рішення. За наявності надійних даних минулих обстежень здійснюють оптимальне розміщення за кожною ознакою окремо. Якщо оптимальні обсяги вибірок відрізняються, задовільне компромісне розміщення забезпечують середні з цих обсягів.

Розміщення вибірки за областями дослідження

Розглянуті способи розміщення вибірки за стратами мали за мету забезпечити надійність оцінок таких загальних параметрів сукупності як середня \bar{x} або частка p . Інша функція поділу сукупності пов'язана з пошуком надійних сепаратних оцінок для кожної складової (\bar{x}_j та p_j). Наприклад, при обстеженні домогосподарств оцінюються середні доходи та витрати у групах сімей за кількістю дітей (0, 1, 2, ...), для домовласників і орендарів житла тощо. Аби вирізняти функції поділу сукупності й уникнути плутанини, для таких складових Статистичним Бюро ООН² рекомендовано термін *область дослідження*. Іноді виокремлені складові виступають в обох ролях: як страти і як області дослідження.

² Підкомісією зі статистичної вибірки.

Якщо мета вибіркового обстеження — порівняти між собою різні області дослідження, то оцінюється не загальна середня ознака \bar{x} , а різниця між середніми цих областей. Якщо областей лише дві, необхідно розмістити вибірку між ними так, щоби стандартна похибка різниці оцінок середніх $(\bar{x}_1 - \bar{x}_2)$ була мінімальною. Скажімо, необхідно визначити, яка група студентів університету (хлопці чи дівчата) в середньому більше відвідують спортивні секції, якщо серед студентів дівчат удвічі більше, ніж хлопців. Стандартна похибка різниці середніх без поправки на скінченність сукупності дорівнює:

$$\mu_{(\bar{x}_1 - \bar{x}_2)} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}.$$

Якщо функція витрат на обстеження лінійна і має вигляд:

$$C = c_0 + c_1 n_1 + c_2 n_2,$$

то стандартна похибка буде мінімальною за такого розміщення вибірки:

$$n_1 = \frac{ns_1/\sqrt{c_1}}{s_1/\sqrt{c_1} + s_2/\sqrt{c_2}}; \quad n_2 = \frac{ns_2/\sqrt{c_2}}{s_1/\sqrt{c_1} + s_2/\sqrt{c_2}}.$$

Припускаючи, що витрати на обстеження для обох областей однакові, стандартна похибка різниці буде мінімальною в тому разі, коли розподіл пропорційний щодо стандартних відхилень s_j :

$$n_1 = \frac{ns_1}{s_1 + s_2}; \quad n_2 = \frac{ns_2}{s_1 + s_2}.$$

Коли областей більше двох ($m > 2$), то оптимальне розміщення залежить від бажаного рівня точності при порівнянні різних областей — μ_{jh} . Наприклад, можна мінімізувати витрати за наявності $m(m-1)/2$ обмежень у вигляді $\mu(x_j - x_j) \leq \mu_{jh}$. Простіший метод передбачає мінімізацію середнього значення стандартної похибки різниці оцінок для всіх $m(m-1)/2$ пар областей:

$$\mu^2 = \frac{2}{m} \left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} + \dots + \frac{s_m^2}{n_m} \right].$$

Якщо області дослідження не вдалося ідентифікувати до формування вибірки, можна їх виокремити за даними вибірки. При достатньому обсязі n_j точність оцінок \bar{x}_j буде адекватною поставленій меті.

Часто області дослідження представлені в усіх апріорі виокремлених стратах. Наприклад, у соціологічному опитуванні при стратифікації за географічною ознакою необхідно знайти оцінки для населення загалом, нарізно для чоловіків і жінок, для різних вікових груп, для сімейних і однаків тощо. Квадрат стандартної похибки середньої для j -ої області дослідження із k -ої страти визначається за формулою:

$$\mu_{kj}^2 = \sum_k \frac{N^2 s^2}{n_{kj}} \left(1 - \frac{n_{kj}}{N_{kj}}\right),$$

де N_{kj} і n_{kj} — кількість одиниць в j -ій області k -ої страти в генеральній і вибірковій сукупностях відповідно.

Стратифікована вибірка вимагає більших витрат на організацію дослідження порівняно з власне випадковим добором. Тому важливо знати, який вигрaш у точності вибіркових оцінок можна отримати за однакового обсягу вибірки, незважаючи на додаткові витрати.

Стратифікація у практиці державних вибіркових обстежень

Класичним прикладом стратифікованої вибірки є структурне вибіркове обстеження діяльності малих підприємств. Мета обстеження — отримати інформацію про соціально-економічний стан і господарську діяльність малих підприємств. Цільові ознаки: обсяг реалізованої продукції (робіт, послуг) і кількість найманих працівників. В Україні обстеження малих підприємств проводиться органами державної статистики на постійній основі з річною періодичністю і базується на загальноприйнятих міжнародних стандартах, відповідно до яких розроблена Методика формування вибіркової сукупності звітних одиниць у структурному обстеженні малих підприємств [Методика формування, 2010].

Генеральна сукупність обстеження формується суто зі звітних одиниць, ідентифікованих в Єдиному державному реєстрі підп-

приємств і організацій України. Основу вибірки становить сукупність малих активних підприємств з кількістю найманих працівників 10–49 осіб. Мале підприємство є одиницею спостереження й одиницею добору, а тому основа вибірки збігається із генеральною сукупністю. Узагальнені характеристики вибіркового структурного обстеження малих підприємств наведено в табл. 7.5.

Таблиця 7.5

**УЗАГАЛЬНЕНІ ХАРАКТЕРИСТИКИ ВИБІРКОВОГО
СТРУКТУРНОГО ОБСТЕЖЕННЯ МАЛИХ ПІДПРИЄМСТВ**

Генеральна сукупність (основа вибірки)	Сукупність усіх малих активних підприємств: – за видами економічної діяльності (КВЕД): секції від <i>A</i> до <i>S</i> , за винятком секції <i>O</i> , та видів діяльності 51.22 секції <i>H</i> ; 64.11, 64.19 секції <i>K</i> ; 94.11.0- 91.99 секції <i>S</i> ; – за організаційно-правовими формами господарювання (КОПФГ): 120-193, 200-270, 300, 310, 330, 340, 500-590, 900, 915-940; – за географічним охопленням: економіка України загалом і в розрізі регіонів (за КОАТУУ)
Джерело інформації для формування основи вибірки	Статистичний реєстр підприємств
Вид вибірки	Стратифікована
Ознаки стратифікації	Основний вид економічної діяльності (згідно з КВЕД); Розмір підприємства, який визначають за кількістю найманих працівників, утворюючи групи: 1) 0–4 особи; 2) 5–9 осіб; 3) 10–19 осіб; 4) 20–50 осіб. Територіальна належність підприємства (шість регіонально-локальних груп)
Розміщення вибірки	Оптимальне розміщення Неймана: $n_j = \frac{N_j s_j}{\sum_1^m N_j s_j} n$
Обсяг і структура вибірки	Обсяг вибірки не має перевищувати 20 % генеральної сукупності. Структура вибірки: підприємства тих видів економічної діяльності, які обстежують суцільно (100 %); нетипові підприємства та малі страти (100 %); 3 % сукупності незначних підприємств; 10 % сукупності новостворених підприємств; приблизно 25 % підприємств основної сукупності.

Оскільки генеральна сукупність малих підприємств за цільовими ознаками обстеження *a priori* неоднорідна, отримання репрезентативної вибірки можливе за умови стратифікації основи вибірки, поділу її на більш однорідні групи (страти), з яких добирається певна кількість одиниць для безпосереднього обстеження.

Стратифікацію здійснюють окремо для основної сукупності, незначних і новостворених підприємств. Стосовно основної сукупності малих підприємств (за винятком нетипових підприємств і малих страт), стратифікацію генеральної сукупності проводять за такими параметрами:

- вид економічної діяльності за секціями і підсекціями КВЕД;
- розмір підприємств за кількістю найманих працівників;
- територіальна належність підприємства.

Новостворені і незначні підприємства стратифікуються за основним видом економічної діяльності.

Дизайн вибірки визначається такими особливостями її формування, як: стратифікація генеральної сукупності; виявлення нетипових одиниць у стратах; об'єднання малих страт. Послідовність процедур формування вибіркової сукупності малих підприємств ілюструє рис. 7.6.

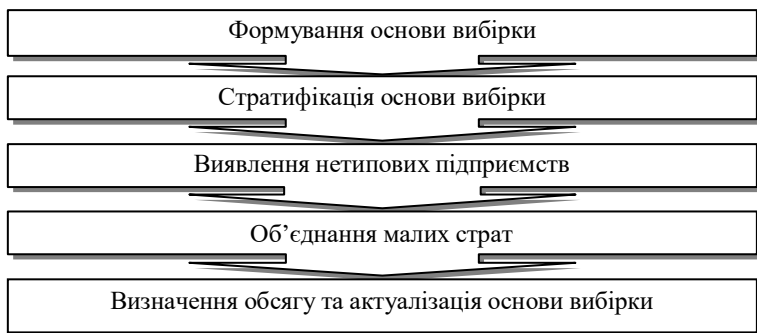


Рис. 7.6. Послідовність процедур формування вибіркової сукупності малих підприємств

За видом економічної діяльності вирізняють 60 груп, за кількістю найманих працівників — 4 групи. Враховуючи те, що стратифікація за регіонами в поєднанні з видом економічної діяльності дає значну кількість дрібних і неоднорідних страт, здійснюють об'єднання регіонів у 6 регіонально-локальних груп. У процесі

формуванні цих груп враховують структуру сукупності за обсягом реалізованої продукції, за видами економічної діяльності, за регіонами. В процесі стратифікації кожному підприємству присвоюють код страти, до якої воно належить, за такою структурою: $XX(XXX)YZ$, де $XX(XXX)$ — основний вид економічної діяльності (за секціями і підсекціями КВЕД);

Y — код локально-регіональної групи;

Z — код групи за розміром підприємства.

Наступна процедура спрямована на виявлення нетипових підприємств та розміщення їх в окремій страті для подальшого обстеження на суцільній основі. До нетипових відносять підприємства, для яких значення обсягу реалізованої продукції потрапляє за межі інтервалу:

$$|y_{jh} - \bar{y}_h| < 3s_j,$$

де y_{jh} — значення обсягу реалізованої продукції для h -ої одиниці в j -ій страті;

\bar{y}_j і s_j — середнє значення і стандартне відхилення обсягу реалізованої продукції в j -ій страті, відповідно.

Для виявлення нетипових підприємств в j -ій страті їх розміщують у порядку зменшення обсягу реалізованої продукції. Наприклад, основні статистичні показники в умовній страті XXX , яка налічує 345 підприємств такі: середній обсяг реалізованої продукції — 1529,7 тис. грн, стандартне відхилення — 1705,21, межа нетиповості — $|y_{jh} - \bar{y}_j| < 3 \cdot 1705,21 = 5115,63$. Підприємства, обсяг реалізації продукції яких не задовольняє цю умову, розглядаються як нетипові (табл. 7.7).

Таблиця 7.7

РЕЗУЛЬТАТИ ПЕРЕВІРКИ УМОВИ НЕТИПОВОСТІ МАЛИХ ПІДПРИЄМСТВ (МП) ЗА ОБСЯГОМ РЕАЛІЗОВАНОЇ ПРОДУКЦІЇ У СТРАТІ XXX

№ з/п	\bar{y}_{jh}	$ \bar{y}_{jh} - \bar{y}_j $	Типові МП $ \bar{y}_{jh} - \bar{y}_j < 3\bar{y}_j$	Нетипові МП $ \bar{y}_{jh} - \bar{y}_j > 3\bar{y}_j$
1	13011,7	11482,0		11482,0 > 5115,6
2	7720,4	6190,7		6190,7 > 5115,6
3	6415,2	4885,5	4885,5 < 5115,6	
4	5942,6	4412,9	4412,9 < 5115,6	

У наведеному прикладі до нетипових належать підприємства № 1–2. Виключення нетипових підприємств (з подальшим об'єднанням їх в окрему страту) сприяє зменшенню похибок вибірки і підвищенню надійності вибіркових оцінок за кожним видом економічної діяльності.

Страти, обсяг яких менше 10 одиниць, вважаються малими. Оскільки малі страти обстежуються суцільно і тим самим вичерпують резерв вибірки, дизайном вибірки передбачено зменшення їхньої кількості шляхом об'єднання з близькими їм стратами. Об'єднання проводять як за локально-регіональними групами, так і за розміром підприємств у межах одного виду економічної діяльності за умови, що в об'єднаних стратах відсутні нетипові підприємства, а ідентичність середніх підтверджує t -критерій для незалежних вибірок. Отже, об'єднання малих страт виконують як для зменшення кількості страт, так і для зниження похибки вибірки.

Якщо з економічної точки зору можливо об'єднати регіони, а t -критерій підтверджує ідентичність середніх, приймається рішення про об'єднання двох регіонів в один. Об'єднання страт проводять поступово. Спершу об'єднуються страти в межах одного виду економічної діяльності та локально-регіональної групи (тобто ті, що відрізняються лише групою за розміром підприємств), потім ті страти, що відрізняються територіальною належністю (в межах одного виду економічної діяльності та групи за розміром). Тобто для страт, закодованих як $XX(XXX)YZ$, об'єднання відбувається спочатку на підставі Z , потім — на підставі Y . Після об'єднання малих страт проводять повторну агрегацію основної сукупності з повторним розрахунком статистичних показників по кожній страті. На цьому етапі відбувається остаточне виявлення нетипових одиниць. За результатами досліджень з урахуванням оцінок точності параметрів загальний обсяг вибіркової сукупності малих підприємств з урахуванням оцінок точності параметрів не має перевищувати 20 % обсягу генеральної сукупності (див. табл. 7.5).

Розподіл вибірки за стратами здійснюють за алгоритмом оптимального розміщення Неймана:

$$n_j = \frac{N_j s_j}{\sum_1^m N_j s_j} n$$

Обсяг вибірки для основної сукупності підприємств поділяється на дві частини: 90 % — для розподілу за стратами на основі визначеного способу розміщення, 10 % — резерв, який використовують для коригування розмірів вибіркової страти. Ця процедура необхідна для впорядкування кількості підприємств у стратах у разі неотримання відповідей в процесі вибіркового обстеження, що є суттєвим для малочисельної страти при поширенні її даних на генеральну сукупність. Розподілення резерву відбувається після оптимального розміщення вибірки за стратами.

Вибіркова сукупність формується з кожної вищевказаної складової основи вибірки окремо за алгоритмом систематичного добору. Розмір вибірки в страті коригується (збільшується чи зменшується) на відсоток невідповідей (за результатами попередньо проведеного обстеження).

Передбачається, що за результатами обстеження малих підприємств рівень надійності оцінювання основних показників відповідатиме таким вимогам:

- на національному рівні гранична похибка вибірки показників не перевищуватиме 5 %;
- за сукупністю звітних одиниць у кожному з регіонів та за кожним із видів економічної діяльності — 10 %;
- за сукупністю звітних одиниць за видами економічної діяльності в межах регіонів та за адміністративно-територіальними одиницями — 15 %.

Зазначені критерії надійності статистичної інформації відповідають рекомендаціям експертів та міжнародній практиці.

Вправи та запитання для самоконтролю

1. Поясніть сутність стратифікованої вибірки і сфери її використання.
2. За яких умов стратифікована вибірка забезпечує вигреш у точності?
3. Як оцінити ефект стратифікації?
4. Оцініть відносну точність зв'язку між ознакою стратифікованої вибірки порівняно з простою, якщо щільність зв'язку між ознакою стратифікації і ознакою, що вивчається, характеризується коефіцієнтом кореляції $r = 0,6$.
5. За даними 19 %-го вибіркового обстеження домогосподарств витрати часу на домашні справи значною мірою залежать від наявності неповнолітніх дітей:

Кількість дітей	Кількість домогосподарств	Витрати часу на домашні справи, год. на тиждень	Стандартне відхилення витрат часу, s_j
Без дітей	200	26	10
Одна дитина	120	30	20
Дві і більше	80	35	15

Визначте:

а) середній тижневий розмір витрат часу на домашні справи для всієї сукупності обстежених домогосподарств та довірчі межі для середньої з імовірністю 0,954;

б) з якою імовірністю можна стверджувати, що тижневий обсяг витрат часу на домашні справи не перевищує 40 годин?

6. Проведено вибіркоче обстеження втрат зерна озимої пшениці через несвоєчасне збирання врожаю. Кількість пробних ділянок визначалася пропорційно посіву відповідного сорту пшениці. Результати обстеження такі:

Сорт пшениці	Кількість пробних ділянок	Втрати зерна, ц/га	Дисперсія втрат зерна
А	10	2	6,4
В	6	7	7,8
С	4	9	10,3

Визначте:

а) середні втрати зерна у розрахунку на одну пробну ділянку та довірчі межі середніх втрат з імовірністю 0,954;

б) мінімально достатній обсяг вибірки, за якого похибка вибірки з тією ж імовірністю не перевищить 1 ц/га.

7. Дирекція мережі супермаркетів провела вибіркоче опитування працівників двох регіонів щодо підтримки ними нової маркетингової політики:

Регіон	Опитано	Підтримують, %
А	24	70
Б	16	80

Визначте середній рівень підтримки нової маркетингової політики і граничну похибку вибірки цього показника з імовірністю 0,954.

Оцініть виграш у точності вибіркових оцінок за рахунок стратифікації.

8. Поясніть відмінності між пропорційним і оптимальним розміщеннями стратифікованої вибірки.

9. За даними 1%-го вибіркового обстеження підприємств середнього бізнесу витрати на оренду приміщень становили:

Район розміщення	Кількість підприємств	Середні витрати, грн на місяць за 1 кв. м	Дисперсія витрат
Середмістя	500	750	3695
Околиці	700	620	1838

Визначте: а) середні витрати підприємств на оренду приміщень; б) стандартну похибку стратифікованої вибірки середніх витрат на оренду приміщень. Оцініть виграш у точності за рахунок стратифікації.

10. З метою включення міста до міжнародного рейтингу привабливості для туристів проведено обстеження закладів громадського харчування (стратифікація за цінovими категоріями: економ, комфорт, бізнес-клас, преміум) для оцінки середньої вартості обіду. За результатами обстеження загальна дисперсія оцінок становить 1116,8, міжгрупова дисперсія — 834,2. Необхідно оцінити виграш точності стратифікованої вибірки порівняно з простою випадковою.

11. Проведено 10 %-е обстеження зареєстрованих у службі зайнятості безробітних з метою оцінки тривалості пошуку роботи залежно від рівня освіти (вища, середня спеціальна, середня). Обсяг вибірки — 49 осіб. За результатами обстеження загальна середня тривалість пошуку роботи становить 6 тижнів, дисперсія — 5,8. Про залежність тривалості пошуку роботи від рівня освіти безробітних свідчить кореляційне відношення $\eta^2 = 0,72$. Визначте стандартну похибку середньої тривалості пошуку роботи. Оцініть виграш точності стратифікованої вибірки порівняно із власне випадковою.

12. Проектується вибіркoве обстеження підприємств з метою визначення витрат на впровадження інновацій (y %). За результатами попередніх обстежень відомо:

Група підприємств за розділом КВЕД	Кількість підприємств, N_j	Середні витрати, млн грн на рік	Стандартне відхилення, s_j
М 71	250	6,3	4,5
М 72	400	9,5	5,0

За яким способом необхідно розмістити вибіркoву сукупність обсягом $n = 36$ за видами економічної діяльності, щоб оцінки витрат на інновації у кожній групі підприємств були визначені з однаковою точністю?

Оцініть ефект оптимального розміщення.

13. Проектується вибіркoве обстеження підприємств хімічної промисловості для визначення середнього обсягу шкідливих викидів.

Оснoву вибірки становить єдиний державний реєстр підприємств та організацій. За умови стратифікованої вибірки очікуються витрати на обстеження:

Страта — масштаб підприємства	Вага страти, %	Стандартне відхилення, s_j	Одиничні витрати на обстеження, грн
Великі	40	15	100
Малі	60	10	80

Визначте: обсяги вибірки за стратами, які б мінімізували загальні витрати на обстеження при заданій точності результатів; обсяг вибірки, який би гарантував у разі оптимального розміщення стандартну похибку на рівні 10 грн.

14. Проектується вибіркoве обстеження. За умови стратифікованої вибірки очікуються витрати на обстеження в розмірі $\sum c_j n_j$. Попередні оцінки відповідних величин по стратам такі:

Страта	Вага страти, %	Стандартне відхилення, s_j	Одиничні витрати на обстеження, грн
1	60	10	8
2	40	5	3

Визначте обсяги вибірки по стратам, які б мінімізували загальні витрати на обстеження при заданій точності результатів.

15. Проектується вибіркoве дослідження торговельних підприємств однакової спеціалізації у межах міста для визначення рівня рентабельності їхньої діяльності. Оскільки місце розташування впливає на структуру покупців й асортимент товарів на їхні цінові категорії, магазини можна поділити на групи: розташовані у середмісті та розташовані на околиці міста. За умови стратифікованої вибірки плануються такі попередні оцінні витрати на обстеження за стратами:

Страта	Вага, %	Стандартне відхилення, s_j	Одиничні витрати на обстеження, грн
1	45	9	7
2	55	7	4

Необхідно визначити: обсяги вибірки за стратами, які б мінімізували загальні витрати на обстеження, обсяг вибірки, який би гарантував у разі оптимального розміщення стандартну похибку на рівні 5 грн.

16. Передбачається обстеження компаній-виробників продуктивних товарів з метою визначення рівня покриття інформаційними каналами реальних обсягів продажу (y %). Частка добору становить 20 %. Вибіркова сукупність — 15 компаній. За результатами попередніх обстежень відомі дані:

Канал збуту	Кількість компаній, N_j	Стандартне відхилення, s_j
Large	12	0,7
Medium	28	1,3
Small	35	1,6

Здійсніть пропорційне та оптимальне розміщення вибірки. Оцініть ефект оптимального розміщення.

17. Планується вибіркове обстеження домогосподарств у місті з метою оцінювання середньої вартості нерухомого майна на одне домогосподарство. Домогосподарства поділені на дві групи: 1 — мешкають у багатоквартирних будинках ($N_1 = 20000$), 2 — мають власні будинки ($N_2 = 4000$). Передбачається, що вартість майна на одне домогосподарство у другій групі приблизно в дев'ять разів більше порівняно з першою і стандартне відхилення може бути пропорційним квадратному кореню з середнього значення по групі.

Як би Ви розмістили вибірку обсягом $n = 1000$ домогосподарств між цими групами? Оцініть відносну точність Вашого варіанта.

Рекомендована література

1. *Гладун О.М.* Вибіркові обстеження населення: методологія, методика, практика : Монографія. — Ніжин: ТОВ «Видавництво «Аспект-Поліграф», 2008. — 344 с.

2. *Гончар О.В.* Забезпечення якості статистичної інформації: методологія та організація : Монографія. — К. : ВПД «Формат», 2011. — 372 с.

3. *Єріна А.М.* Організація вибірових обстежень : Навчальний посібник. — К. : КНЕУ, 2003. — 127 с.

4. Методика формування вибіркової сукупності звітних одиниць у структурному обстеженні малих підприємств / Державна служба статистики України. URL: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2010/536/metod.htm

5. Статистика підприємств : Навчальний посібник / *За ред. С.О. Матковського.* — К. : Алерта, 2017. — 560 с.

Розділ 8

БАГАТОСТУПЕНЕВІ ВИБІРКОВІ ОБСТЕЖЕННЯ

Схеми багатоступеневого (багатостадійного) добору
Оптимальне співвідношення обсягів вибірки на різних
ступенях добору
Організація вибіркових обстежень населення (домого-
сподарств)

Після вивчення цього розділу студент має

- *знати*: принципи формування багатоступеневої вибірки, які забезпечать економію витрат на обстеження без істотного зниження точності результатів, на підставі даних з різних джерел інформації; способи добору первинних і локальних територіальних одиниць;
- *уміти*: формувати основу вибірки на кожному ступені добору; добирати несамопрезентативні локальні одиниці вибірки та територіальні одиниці за принципом географічного серпантину; оцінювати точність результатів і дизайн-ефект багатоступеневої вибірки.

Схеми багатоступеневого (багатостадійного) добору

Багатоступенева вибірка досить поширена; її вважають зручним і гнучким способом організації вибіркових обстежень, що здатен покращити баланс між точністю результатів і витратами на обстеження.

Характерною рисою багатоступеневої вибірки є те, що на кожному ступені добору виокремлюють різні проміжні об'єкти репрезентації, які співвідносяться за принципом «матрьошки». На першому ступені добираються укрупнені одиниці, переважно це природні формування (географічні регіони, адміністративні одиниці тощо); на другому — серед дібраних укрупнених одиниць добираються менші за обсягом частини, і так доти, доки не будуть дібрані ті окремі, за висловом М.К. Дружиніна, «елемен-

тарні одиниці», які безпосередньо підлягають обстеженню [Дружинін, 1977].

Ступенів добору може бути два, три і більше. На кожному з них формується своя основа вибірки об'єктів репрезентації, в яких концентруються одиниці спостереження. Повнота охоплення основи й схема добору об'єктів репрезентації для обстеження на різних ступенях різняться. Практично всі відомі великомасштабні соціально-економічні обстеження здійснено на основі багатоступеневої вибірки, в яких одиницями першого ступеня добору є переважно адміністративні одиниці, населені пункти або виокремлені страти (установи, навчальні заклади, лікарські дільниці тощо). Упорядкування списку порівняно невеликої кількості укрупнених одиниць набагато простіше, ніж списку великої сукупності одиниць спостереження. Крім того, при багатоступеневому доборі, коли об'єктами репрезентації першого ступеня є географічні регіони чи адміністративні одиниці, скорочуються територіальні кордони обстеження, завдяки чому багатоступеневий добір порівняно з одноступеневим виявляється більш економічним.

Багатоступеневий добір поєднує принципи кластерної й систематичної вибірок. Кластером (гніздом) вважають той проміжний об'єкт дослідження, який добирається на кожному ступені і виступає первинною сукупністю для наступного ступеня добору. На відміну від одноступеневого кластерного добору в багатоступеневій вибірці обстеженню підлягають не всі одиниці кластера, а лише частина, дібрана за принципом випадкової вибірки.

Логіку багатоступеневої вибірки розглянемо на прикладі двоступеневого добору, який схематично подано на рис. 8.1. Припустімо, сукупність містить $N = 25$ первинних одиниць (гнізд) по $M_0 = 4$ вторинні одиниці у кожному з них. При формуванні вибіркової сукупності на першому ступені добирається $n = 5$ первинних одиниць (гнізд), на другому ступені — по $m_0 = 2$ вторинних одиниць у кожному дібраному гнізді. Обсяг вибіркової сукупності становить $(n \times m_0) = 5 \times 2 = 10$ одиниць спостереження.

При двоступеневому доборі похибка вибірки має два джерела:

- варіацію між первинними одиницями (гніздами), мірою якої є міжгніздова дисперсія s_B^2 ;

- варіацію між вторинними одиницями всередині гнізд, яка вимірюється внутрішньогніздовою дисперсією s_w^2 ;

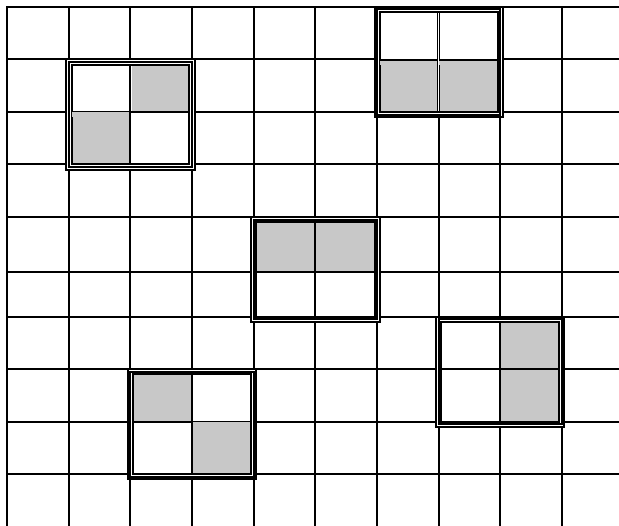


Рис. 8.1. Схема простого двоступеневого добору

Якщо проігнорувати поправку на скінченність сукупності, похибка вибірки середньої в розрахунку на одиницю спостереження (вторинну одиницю) розраховується за формулою:

$$\mu^2(\bar{x}') = \frac{s_w^2}{nm_0} + \frac{\mathbb{E}}{n}.$$

Як видно з формули, збільшення кількості дібраних вторинних одиниць m_0 зменшує лише одну складову похибки $\frac{\mathbb{E}}{n}$, тоді як збільшення кількості первинних одиниць n при $m_0 = \text{const}$ зменшує обидві її складові.

Розрахунок оцінок вказаних вище дисперсії зручно вести в таблиці дисперсійного аналізу результатів вибірки (табл. 8.1). Позначимо середні квадрати відхилень: для всіх джерел варіації — s^2 , внутрішньогрупової варіації — s_w^2 , для міжгрупової варіації — s_B^2 . Взаємозв'язок між цими складовими дисперсії такий

$$s^2 = \frac{n(m_0 - 1)}{nm_0 - 1} s_w^2 + \frac{n - 1}{nm_0 - 1} s_B^2.$$

Наприклад, проведено соціологічне обстеження, завданням якого було в'яснити середню кількість звернень мешканців району міста за лікарською допомогою впродовж року. Для обстеження відібрано шість будинків (гнізд). У кожному будинку до вибірки потрапили мешканці п'яти квартир. За результатами опитувань у середньому за рік на одну квартиру припадало 11 звернень до лікарів. Загальна сума квадратів відхилень для 30 квартир становила 650, для шести будинків — 230.

Таблиця 8.1

ДИСПЕРСІЙНИЙ АНАЛІЗ ЗВЕРНЕНЬ ЗА ЛІКАРСЬКОЮ ДОПОМОГОЮ

Джерело варіації	Сума квадратів відхилень, S^2	Число ступенів свободи, df	Середній квадрат відхилень, $s^2 = S^2 / df$
Усі	650	29	22,41
Будинки	230	5	46,0
Квартири	420	24	17,5

За даними таблиці перевіримо взаємозв'язок середніх квадратів відхилень й оцінимо компоненти дисперсії:

$$s^2 = \frac{6 \cdot (5-1)}{6 \cdot 5 - 1} 17,5 + \frac{6-1}{6 \cdot 5 - 1} 46 = 22,41.$$

Компоненти дисперсії:

$$s_W^2 = 17,5; \quad s_B^2 = (46,0 - 17,5)/5 = 5,7.$$

$$\text{Звідси } \mu^2 = \frac{5,7}{6} + \frac{17,5}{6 \times 5} = 1,53, \text{ а } \mu = \sqrt{1,53} = 1,24.$$

Альтернативним методом розрахунку похибки вибірки μ^2 є використання коефіцієнта внутрішньогніздової кореляції δ :

$$\mu^2 = \frac{s^2}{nm_0} [1 + (m_0 - 1)\delta],$$

$$\text{де } \delta = \frac{s^2 - s^2}{m_0 s^2} \text{ або приблизно } \delta \cong \frac{s_B^2}{s^2}.$$

Коефіцієнт кореляції δ вимірює зв'язок між парами спостережень в одному й тому ж гнізді. Значення його змінюється в межах $[-(1/m_0 - 1) \leq \delta \leq +1]$. Якщо гнізда є «випадковими утвореннями», то $\delta = 0$, а похибка гніздової вибірки дорівнює похибці простої випадкової вибірки. За даними табл. 8.1:

$$\delta = \frac{46 - 17,5}{5 \cdot 22,41} = \frac{28,5}{112,05} = 0,254, \text{ або } \delta = \frac{s_B^2}{s^2} = \frac{5,7}{22,41} = 0,254.$$

$$\text{Тоді } \mu^2 = \frac{22,41}{6 \times 5} [1 + (5 - 1) \cdot 0,254] = 1,51.$$

У розглянутому прикладі первинні вибіркові одиниці однакового розміру ($m_0 = 5$), і при формуванні вибірки на обох ступенях можна застосувати добір з однаковою ймовірністю. Таку схему вибірки легко поширити на три і більше ступенів добору (за умови однакового розміру вибіркових одиниць).

Під час організації багатоступеневої вибірки в соціально-економічних дослідженнях часто стикаються із ситуацією, коли проміжні одиниці добору неоднакові за розміром. Іноді за допомогою стратифікації, якщо ознакою стратифікації є розмір одиниць, можна виокремити кластери (гнізда) однакового чи приблизно однакового розміру. Але в переважній більшості реальні кластери неоднакові за розміром, що вимагає застосування інших схем добору. Найпоширенішою серед них є систематичний спосіб добору з імовірностями, пропорційними розміру проміжних вибіркових одиниць (*probability proportional to size*, PPS). Головна перевага такого способу добору полягає в тому, що для багатоступеневих вибірок він дозволяє забезпечити однакоку загальну ймовірність добору одиниць вибірки останнього ступеню за умови рівномірного розподілу їх за первинними кластерами різного обсягу.

Як характеристику розміру кластера використовують або кількість одиниць спостереження, або будь-який інший показник, пов'язаний із сумарним значенням одиниць спостереження. Наприклад, розмір компанії можна схарактеризувати вартістю капіталу чи активів, чисельністю зайнятих чи обсягом виробництва. Аналогічно можна обґрунтувати різні характеристики розміру банку, ресторану, лікарні тощо. Розмір територіальної одиниці

характеризує чисельність населення, кількість домогосподарств або розмір земельної ділянки домогосподарств.

Формування багатоступеневої вибірки відносять до складних форм роботи з досліджуванним об'єктом. Цей процес охоплює кілька відносно самостійних етапів, на кожному з яких використовують той чи інший із раніше розглянутих методів. Схема добору з імовірностями, пропорційними розміру кластера, реалізується в кілька етапів по аналогії із систематичним добором (рис. 8.2).

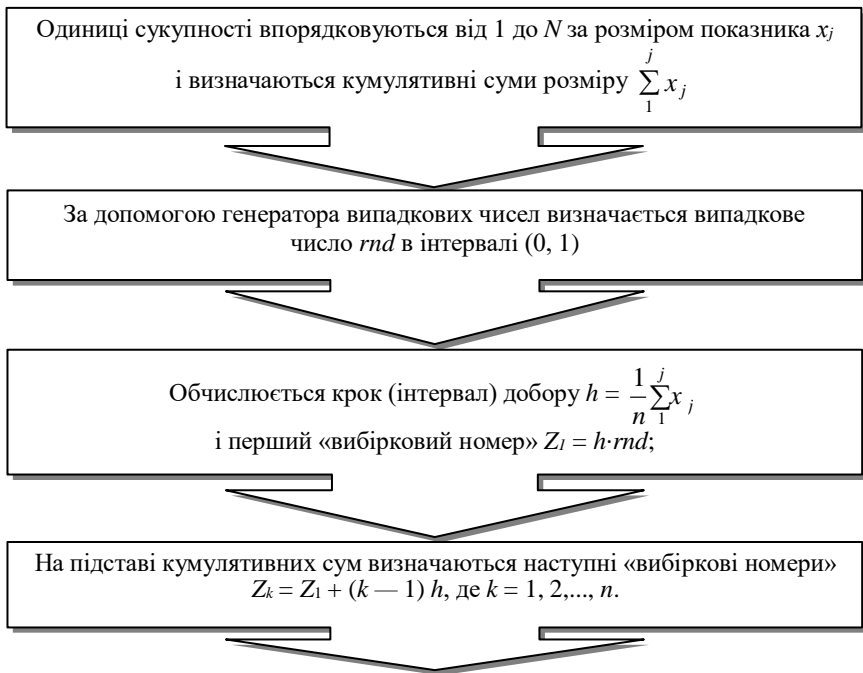


Рис. 8.2. Послідовність формування багатоступеневої вибірки за схемою добору з імовірностями, пропорційними розміру кластера (PPS)

Дібраними вважаються ті одиниці списку, на які припадають «вибіркові номери». Методика добору вибірових одиниць з імовірністю, пропорційною їхньому розміру, ілюструється в табл. 8.2.

Таблиця 8.2

ПОРЯДОК ДОБОРУ, ПРОПОРЦІЙНОГО РОЗМІРУ ВИБІРКОВИХ ОДИНИЦЬ

Одиниця списку	Розмір одиниці, x_j	Кумулятивна сума розміру, $\sum_1^j x_j$	Вибіркові номери, Z_k	Дібрані одиниці списку, j
1	6	6	4	1
2	5	11		
3	4	15	14	3
4	3	18		
5	2	20		

Проміжні вибіркові одиниці розміщені у списку в порядку зменшення розміру. Загальна сума значень ознаки (розміру) становить $\sum_1^n x_j = 20$, вибіркова сукупність $n = 2$, крок інтервалу — $(20 : 2) = 10$, випадкове число $rnd = 0,4$. Звідси, перший вибірковий номер дорівнює $Z_1 = 10 \times 0,4 = 4$, другий — $Z_2 = (4 + 10) = 14$. Ці вибіркові номери належать одиницям списку № 1 і № 3, які й формують вибіркову сукупність.

Розмір одиниць сукупності можна подати відносною величиною (часткою), тоді $\sum_1^n x_j = 1$, крок інтервалу $h = 1/n$, випадкове число визначають у межах від 0 до $1/n$.

Отже, при доборі, пропорційному розміру вибіркових одиниць, основою вибірки виступають кумулятивні суми того показника, який характеризує розмір одиниці списку. Використання такої схеми формування вибірки вважається доцільним у тих випадках, коли одиницями списку є регіони, адміністративні одиниці, інші групування на основі неперервної ознаки.

Точність оцінювання середніх у розрахунку на вторинну одиницю, тобто одиницю спостереження, залежить від співвідношення розмірів первинних одиниць. У табл. 8.3 наведено дані, запозичені в [Кокрен, 1976: с. 318]. У списку три первинні одиниці ($n = 3$) різного розміру. Для характеристики розміру використано кількість вторинних одиниць (M_j) і значення показника розміру x_j , за яким для кожної одиниці списку визначено середні величини (\bar{x}_j) і дисперсії (s_j^2). Загальна середня показника x у розрахунку на вторинну одиницю становить $\bar{x}' = 33 : 12 = 2,75$.

Оцінити точність середньої можна на основі обох характеристик розміру. Формули оцінок точності для зручності подано у вигляді квадратів похибок вибірки μ^2 :

– при доборі, пропорційному кількості вторинних одиниць M_j / M :

$$\mu^2(x'') = \frac{1}{M} \left[\sum_1^N (M_j - m_j) \frac{s_j^2}{m_j} + \sum_1^N M_j (\bar{x}_j - \bar{x}'')^2 \right]$$

– при доборі, пропорційному розміру первинних одиниць за ознакою x :

$$\mu^2(x''') = \frac{1}{M^2} \left[\sum_1^N \frac{M_j(M_j - m_j)}{A_j} \cdot \frac{s_j^2}{m_j} + \sum_1^N A_j \left(\frac{M_j \bar{x}_j}{A_j} - M \bar{x}''' \right)^2 \right]$$

Таблиця 8.3

ХАРАКТЕРИСТИКИ РОЗМІРУ І ВАРІАЦІЇ ПРОМІЖНИХ ОДИНИЦЬ ДВОСТУПЕНЕВОГО ДОБОРУ

№ з/п одиниці	x_{ij}	M_j	M_j / M	$\sum_1^n x_j$	\bar{x}_j	s_j^2
1	0; 1	2	0,17	1	0,5	0,500
2	1; 2; 2; 3	4	0,33	8	2,0	0,667
3	3; 3; 4; 4; 5; 5	6	0,50	24	4,0	0,800
Разом	X	12	1,0	33	2,75	X

Розрахунок складових квадрата похибки вибірки для першого варіанта міститься в табл. 8.4. За даними таблиці:

$$\mu^2(x'') = \frac{2,267}{12} + \frac{21,750}{12} = 0,189 + 1,813 = 2,002,$$

тобто визначальною складовою похибки є варіація між первинними одиницями (1,813), тоді як варіація всередині цих одиниць майже у 10 разів менша (0,189).

Другий варіант оцінювання точності середньої застосовують у тих випадках, коли відсутня інформація щодо кількості одиниць спостереження в первинних одиницях. Тоді розміри первинних одиниць (гнізд) характеризуються відносними величинами — вагами A_j , сума яких дорівнює одиниці. Чим точніше ваги A_j ві-

дображають розміри первинних одиниць, тим точнішою буде оцінка середньої \bar{x} за результатами вибірки.

Таблиця 8.4

**ДО РОЗРАХУНКУ ПОХИБКИ ВИБІРКИ СЕРЕДНЬОЇ ПРИ ДОБОРІ,
ПРОПОРЦІЙНОМУ КІЬКОСТІ ВТОРИННИХ ОДИНИЦЬ**

Одиниця	$M_j - m_j$ ($m_j = 2$)	$\frac{s_j^2}{m_j}(M_j - m_j)$	$\bar{x}_j - \bar{x}$ ($\bar{x} = 2,75$)	$(\bar{x}_j - \bar{x})^2 M_j$
1	0	0	-2,25	10,125
2	2	0,667	-0,75	2,250
3	4	1,600	+1,25	9,375
Разом	X	2,267	X	21,750

Розрахунок складових квадрата похибки вибірки для середньої при доборі, пропорційному розміру первинних одиниць за ознакою x , наведено в табл. 8.5. За даними таблиці маємо:

$$\mu^2(\bar{x}) = \frac{30,67}{144} + \frac{516,0}{144} = 0,213 + 3,583 = 3,796,$$

тобто і за цим варіантом визначальною складовою похибки є варіація між первинними одиницями, яка в 17 разів перевищує внутрішньогрупову варіацію.

Таблиця 8.5

**ДО РОЗРАХУНКУ ПОХИБКИ ВИБІРКИ СЕРЕДНЬОЇ ПРИ ДОБОРІ,
ПРОПОРЦІЙНОМУ РОЗМІРУ ВТОРИННИХ ОДИНИЦЬ**

Одиниця	A_j	$\frac{M_j(M_j - m_j)}{A_j m_j} s_j^2$	$\frac{M_j \bar{x}_j}{A_j}$	$\frac{M_j \bar{x}_j}{A_j} - M \bar{x}$	$(\frac{M_j \bar{x}_j}{A_j} - M \bar{x})^2 A_j$
1	0,2	0	5	-28	156,8
2	0,4	6,67	20	-13	67,6
3	0,4	24,0	60	+27	291,6
Разом	1,0	30,67	X	X	516,0

Похибка вибірки для середньої за другим варіантом порівняно з першим значно більша. Це пояснюється тим, що ваги A_j не точно відбивають розміри первинних одиниць. Очевидно, в разі вдалого вибору характеристики розміру первинних одиниць точність оцінювання буде вищою.

Оптимальне співвідношення обсягів вибірки на різних ступенях добору

При організації багатоступеневої вибірки постає низка методологічних питань, зокрема, щодо числа ступенів добору і обсягів вибірки на різних його етапах. Якими гніздами оперувати — великими чи дрібними? Добирати велику кількість дрібних гнізд з невеликою нормою добору з кожного гнізда чи орієнтуватися на невелику кількість гнізд з високою нормою добору з кожного з них? Великим гніздам, як відомо, властива більша дисперсія, що за інших однакових умов надає їм перевагу перед дрібними. До того ж використання великого гнізда вимагає менших транспортних витрат.

Ефективне застосування багатоступеневого добору передбачає оптимальні пропорції між розміром гнізд і обсягами вибірки на різних ступенях формування вибіркової сукупності. Враховуючи, що кожен ступінь добору вимагає певних витрат праці та коштів, ефективною буде така пропорція, яка для досягнення тієї самої точності, що й інші, потребує менших зусиль.

Назагал це завдання зводиться до пошуку таких обсягів вибірки на окремих ступенях добору n_1, n_2, n_3, \dots і співвідношень між ними, які забезпечують мінімальну вартість обстеження C за фіксованого значення стандартної похибки, тобто пошуку мінімуму суми (без накладних витрат):

$$C = c_1 n_1 + c_2 n_1 n_2 + c_3 n_1 n_2 n_3 + \dots,$$

де c_1, c_2, c_3 — одиничні витрати на відповідних ступенях добору.

Оскільки n_1, n_2, n_3, \dots є елементами розрахунку стандартних похибок на окремих ступенях і в цілому, то результат буде різнитися залежно від способу добору на окремих ступенях, від варіації ознаки, від загальної кількості одиниць, з якого здійснюється добір.

Розглянемо принцип оптимізації співвідношення кількості одиниць добору з урахуванням вартості обстеження на прикладі двоступеневої вибірки, для якої:

$$C = c_1 n_1 + c_2 n_1 n_2,$$

де n_1 — кількість гнізд;

n_2 — кількість одиниць, що добираються у кожному гнізді;

$n_1 n_2$ — загальна кількість одиниць, дібраних для обстеження;

c_1 — фіксовані витрати на спостереження гнізда незалежно від його розміру;

c_2 — фіксовані витрати на спостереження окремої одиниці всередині гнізда незалежно від його розміру.

Складові функції вартості обстеження, як і складові похибки двоступеневої вибірки, пропорційні кількості первинних одиниць (гнізд) і загальній кількості одиниць обстеження:

$$\mu^2 = \frac{s_B^2}{n_1} + \frac{s_W^2}{n_1 n_2}.$$

Звідси можна визначити оптимальну кількість гнізд n_1 , обстеження яких за наявної варіації ознаки забезпечить бажану точність результатів вибірки μ :

$$n_1 = \frac{s_B^2}{\mu^2} + \frac{s_W^2}{\mu^2 n_2}.$$

Співвідношення кількості гнізд n_1 і кількості одиниць у кожному гнізді n_2 показує, що за будь-якого значення n_2 значення n_1 не може бути меншим відношення $\frac{s_B^2}{\mu^2}$.

Вираз значення n_1 підставимо у формулу вартості обстеження C . Найменше значення вартості C буде при:

$$n_2 = \sqrt{\frac{s_W^2}{s_B^2} \cdot \frac{c_1}{c_2}},$$

тобто кількість одиниць n_2 , що добираються у кожному гнізді, залежить від двох величин: перша — співвідношення внутрішньої і міжгніздової варіації, друга — співвідношення фіксованих витрат на обстеження вибіркових одиниць першого і другого ступеня. У тих випадках, коли $s_W^2 > s_B^2$, слід вибирати менше гнізд, але з більшою кількістю одиниць спостереження у кожному з них. Якщо витрати на обстеження окремої одиниці великі порівняно з витратами на добір гнізда, то тоді доцільно добирати більше гнізд невеликого розміру.

Наприклад, сумарні витрати часу на обстеження плануються в обсязі 175,5 год. При цьому на одну одиницю другого ступеню витрати часу 3 хв. або $c_2 = 0,05$, на одну одиницю першого ступеня — $c_1 = 1,25$ год. Нехай $s_w^2 = 64$, а $s_B^2 = 16$.

Визначимо, скільки треба дібрати одиниць першого ступеню n_1 і скільки в кожній з них дібрати одиниць другого ступеню n_2 , щоб з найменшими витратами часу забезпечити точність оцінювання на рівні стандартної похибки $\mu = 0,1$.

Підставивши ці дані у формулу, дістанемо:

$$n_2 = \sqrt{\frac{64}{16} \cdot \frac{1,25}{0,05}} = \sqrt{100} = 10;$$

$$n_1 = \frac{16}{0,1^2} + \frac{64}{0,1^2 \cdot 10} = 100 + 40 = 140.$$

Отже, найменші витрати будуть при доборі 140 одиниць першого ступеню з обстеженням у кожній з них по 10 одиниць.

Оптимальну кількість гнізд і їхній розмір можна визначити, спираючись на загальну дисперсію ознаки s^2 і внутрішньогніздової кореляцію δ . Оскільки

$$\mu^2 = \frac{s^2}{n_1 n_2} [1 + (n_2 - 1)\delta],$$

то оптимальний розмір гнізда становить:

$$n_2 = \sqrt{\frac{c_1}{c_2} \cdot \frac{1 - \delta}{\delta}},$$

де δ — коефіцієнт внутрішньогніздової кореляції, який відбиває співвідношення внутрішньої та міжгніздової варіації.

З урахуванням оптимального розміру гнізд n_2 бажану точність результатів обстеження забезпечить така кількість гнізд n_1 :

$$n_1 = \frac{s^2}{\mu^2} \cdot \frac{1 + (n_2 - 1)\delta}{n_2}.$$

Наприклад, витрати на обстеження гнізда в 4 рази перевищують витрати на обстеження окремої одиниці всередині гнізда.

Загальна дисперсія становить 5,6, коефіцієнт внутрішньогніздової кореляції — 0,2. Визначимо кількість гнізд і їхній розмір за умови, що похибка вибірки не перевищить 0,7.

$$\text{Оптимальний розмір гнізда } n_2 = \sqrt{\frac{4}{1} \cdot \frac{1-0,2}{0,2}} = 4 \text{ одиниці.}$$

$$\text{Кількість гнізд } n_1 = \frac{5,6}{0,7} \cdot \frac{1+(4-1)}{4} = 8 \cdot s_W^2 / s_B^2$$

У практиці вибіркових обстежень оптимальні обсяги вибірки на всіх ступенях добору визначають і для пробних обстежень. У табл. 8.6 наведено схеми пробних обстежень, які забезпечують 90 %-у відносну точність результатів [Кокрен, 1976: с. 306]. За даними таблиці для забезпечення такого рівня точності потрібно не більше 10 первинних одиниць (гнізд) за будь-якого співвідношення як дисперсії, так і одиничних витрат на обстеження. На кількість одиниць обстеження всередині дібраних гнізд істотно впливає співвідношення дисперсії, а вплив співвідношення одиничних витрат маловідчутний.

Таблиця 8.6

ЕКОНОМНІ СХЕМИ ПРОБНИХ ОБСТЕЖЕНЬ

s_W^2 / s_B^2	n_1, n_2	c_1 / c_2				
		< 1	2	4	8	16
1	n_1	7	6	6	5	5
	n_2	3	4	5	6	7
2	n_1	8	7	6	6	5
	n_2	5	7	9	9	13
4	n_1	9	8	8	7	7
	n_2	9	11	12	14	15
8	n_1	10	10	9	9	8
	n_2	14	15	17	18	22
16	n_1	10	10	10	10	8
	n_2	25	27	27	28	37

Розглянута методика пошуку оптимальної кількості проміжних і кінцевих вибірових одиниць поширюється на триступеневий добір. За аналогією зі структурою наведених вище формул маємо:

$$n_3 = \sqrt{\frac{s_2^2}{s_3^2} \cdot \frac{c_2}{c_3}}; \quad n_2 = \sqrt{\frac{s_1^2}{s_2^2} \cdot \frac{c_1}{c_2}}.$$

Оскільки загальний обсяг вибірки можна подати добутком $n = n_1 n_2 n_3$, то кількість одиниць, які добираються на першому ступені становить:

$$n = \frac{n_1 n_2 n_3}{2 \ 3}.$$

Наприклад, необхідно встановити найвигідніші пропорції триступеневого добору із генеральної сукупності $N = 20000$ одиниць, яка поділяється на $N_1 = 10$ вибірових одиниць першого ступеню по $N_2 = 25$ вибірових одиниць другого ступеню, у кожній з них нараховують по $N_3 = 80$ кінцевих вибірових одиниць (одиниць спостереження). Варіація вимірюється дисперсіями: $s_1^2 = 1$, $s_2^2 = 7$, $s_3^2 = 56$. Одиничні витрати часу на обстеження становлять, год.: на першому етапі добору — $c_1 = 128$, на другому $c_2 = 36$, на третьому $c_3 = 2$. Передбачено дібрати разом $n = 120$ одиниць, забезпечивши відносну стандартну похибку на рівні не більше 10 %. За наведеними вище формулами маємо:

$$n_3 = \sqrt{\frac{56}{7} \cdot \frac{36}{2}} = 12; \quad n_2 = \sqrt{\frac{7}{1} \cdot \frac{128}{36}} \approx 5,$$

$$n = \frac{n_1 n_2 n_3}{2 \ 3} = \frac{120}{5 \cdot 12} = 2.$$

Отже, на першому ступені необхідно дібрати 2 вибірові одиниці, із кожної з них на другому ступені добору — по 5 вибірових одиниць, з яких на третьому ступені добирається по 12 одиниць спостереження, тобто $2 \times 5 \times 12 = 120$ одиниць.

Організація базових вибіркових обстежень населення (домогосподарств)

Залежно від мети дослідження і витрат на обстеження класичні схеми формування вибірки адаптуються до особливостей реальних сукупностей. Як ілюстрацію такої адаптації розглянемо методологію організації базових вибіркових обстежень населення (домогосподарств).

Будучи основним джерелом інформації щодо рівня, умов життя та економічної активності населення, вибіркові обстеження домогосподарств досить поширені у світі й у межах окремих країн наймасштабніші. Статистичним бюро ООН узагальнено міжнародний досвід і окреслено певні стандарти щодо формування вибіркової сукупності домогосподарств, змісту програм і організації обстеження. В Україні базові вибіркові обстеження населення (домогосподарств) проводять органи державної статистики. До системи державних вибіркових обстежень населення входять три гармонізовані підсистеми: обстеження умов життя домогосподарств, обстеження населення з питань економічної активності та обстеження сільськогосподарської діяльності домогосподарств у сільській місцевості. Усі обстеження проводяться на окремих мережах респондентів. Програми обстежень ґрунтуються на сучасних методологічних принципах та міжнародних стандартах, що забезпечує якість результатів.

Методологія формування вибірок для проведення вибіркових обстежень населення, основні визначення та концепції, принципи та методологічні засади, що застосовуються в обстеженні, назагал відповідають прийнятим у міжнародній, зокрема в європейській, статистичній практиці вимогам і рекомендаціям. Відповідно до вимоги щодо формування вибірок для обстеження населення доцільним визнано застосування багатоступеневої вибірки: у міських поселеннях — триступеневої, у сільській місцевості — двоступеневої.

Задля забезпечення періодичного оновлення територій для базових обстежень населення термін дії територіальної вибірки визначено п'ять років. Враховуючи методологічні особливості вибіркових обстежень, період проведення кожного з них визначається окремо. Обстеження різняться також очікуваним коефіціє-

єнтом відповідей. Термін дії територіальної вибірки і очікувані рівні відповідей у базових вибіркових обстеженнях населення (домогосподарств) України у циклі 2014–2018 років подано в табл. 8.7 [Методика формування, 2013].

Розроблення дизайну та формування основи вибірки для базових вибіркових обстежень населення (домогосподарств) здійснюється з використанням даних демографічної статистики, перепису населення, погосподарського обліку, картографічних матеріалів.

Вирішальним етапом підготовки та проведення базових вибіркових обстежень населення (домогосподарств) є формування вибіркових сукупностей територій та домогосподарств. Як видно на рис. 8.3, процес формування вибірки для обстеження домогосподарств здійснюється у три етапи, а саме:

- добір адміністративно-територіальних одиниць (первинних територіальних одиниць вибірки, ПТОВ);

- добір локальних територій — інструкторських дільниць у міській місцевості та сільських рад у сільській місцевості (за прийнятою термінологією — вторинних територіальних одиниць вибірки, ВТОВ);

- добір домогосподарств (або осіб).

Таблиця 8.7

**ТЕРМІН ДІЇ І ОЧІКУВАНІ РІВНІ ВІДПОВІДЕЙ У ВИБІРКОВИХ
ОБСТЕЖЕННЯХ НАСЕЛЕННЯ (ДОМОГОСПОДАРСТВ) У 2014–2018 РОКАХ**

Вибіркове обстеження	Період проведення обстеження	Очікуваний рівень відповідей		
		Великі міста	Малі міста	Сільська місцевість
Обстеження економічної активності населення	січень 2014 р. — грудень 2018 р.	80	85	90
Обстеження умов життя домогосподарств	листопад 2013 р. — січень 2019 р.	65	85	90
Обстеження сільськогосподарської діяльності населення	травень 2013 р. — січень 2019 р.	X	X	90

Основою вибірки при доборі ПТОВ і ВТОВ є перелік відповідних територіальних одиниць, при доборі домогосподарств — повний список їхніх адрес у межах відповідної територіальної одиниці. Розглянемо в загальних рисах особливості формування основ триступеневої вибірки обстеження населення (домогосподарств) у циклі 2014–2018 років.

З метою адекватного відображення у вибірці особливостей адміністративно-територіального устрою країни, а також для забезпечення добору домогосподарств з більш однорідних за основними характеристиками груп на початковому етапі формування вибірки здійснюють стратифікацію генеральної сукупності за регіонами (територіальні страти) і кожен регіон поділяється на субстрати за типом поселення: міста з чисельністю населення 100 тис. осіб і більше (великі міста), міста і селища міського типу з чисельністю населення менше 100 тис. осіб (малі міста) та адміністративні райони сільської місцевості (райони регіону).

Стратифікація генеральної сукупності з урахуванням адміністративно-територіального устрою країни забезпечує потребу в інформації за типами населених пунктів (міста, сільська місцевість), а також дає можливість врахувати різні умови роботи статистиків з проведення вибірових обстежень населення.

У міських поселеннях формування основи вибірки та добір ПТОВ здійснюється в межах кожної страти окремо. «Великі міста» добирається з імовірністю $P = 1$, тобто кожне місто страти формує окрему ПТОВ. Добір ПТОВ у стратах «малі міста» і «сільські райони» здійснюється на основі механізму випадкового добору з імовірністю, пропорційною їхньому розміру: малі міста — пропорційно чисельності неінституціонального постійного населення³, сільські райони — пропорційно площі землі. Добір локальних територій способом PPS забезпечує самозваженість вибірок (однакову частку добору по стратах і субстратах) і сприяє ефективному вимірюванню цільових показників.

Залежно від розміру територіальної одиниці та чисельності населення, яку має репрезентувати одна територіальна одиниця, «малі міста» поділяються на саморепрезентативні і несаморе-

³ Із загальної чисельності населення виключено інституціональне населення: військовослужбовців строкової служби, осіб, що перебуває в містах позбавлення волі, осіб, які постійно проживають у будинках-інтернатах, будинках для осіб похилого віку, в гуртожитках (крім сімейних), а також маргінальних верств населення.

презентативні. Поріг саморепрезентативності np визначають на підставі обсягу вибірки n та заданої кількості домогосподарств, що обстежуються одним статистиком у межах територіальної одиниці (навантаження статистика):

$$np = \frac{N}{n} \cdot m \cdot mf,$$

де np — поріг саморепрезентативності;

N та n — обсяги генеральної та вибіркової сукупностей, відповідно;

m — навантаження на одного фахівця з інтерв'ювання;

mf — середній розмір домогосподарства (осіб).

Розглянемо порядок розрахунку порогу саморепрезентативності малих міст умовного регіону. Припустимо, що:

— обсяг генеральної сукупності $N = 175\,102$ домогосподарств;

— обсяг вибіркової сукупності $n = 138$ домогосподарств;

— навантаження статистика $m = 18$ домогосподарств;

— середній розмір домогосподарства (осіб) $mf = 2,45$ особи.

За цих умов поріг саморепрезентативності (np) становить:

$$np = \frac{N}{n} \cdot m \cdot mf = \frac{175102}{138} \cdot 18 \cdot 2,45 \cong 56,0 \text{ тис. осіб.}$$

Таким чином, місто з населенням 56,0 тис. осіб і більше визначається як саморепрезентативне і включається у вибірку з імовірністю $P = 1$.

Добір несаморепрезентативних територіальних одиниць першого ступеню (ПТОВ) здійснюється поетапно (рис. 8.4). Насамперед малі міста впорядковуються за чисельністю населення у порядку зменшення. Кумулятивна сума чисельності населення розглядається як числова схема формування вибіркової сукупності територіальних одиниць. Обсяг вибірки, тобто кількість міст і поселень міського типу, які добираються для подальшого обстеження, визначається n діленням загальної чисельності населення у відповідній страті $\sum_1^n x_j$ на поріг саморепрезентативності. Нехай загальна чисельність населення в страті становить $\sum_1^n x_j = 430,0$ тис. осіб, тоді обсяг територіальної вибірки — $(430,0 : 56,0) \cong 7$ одиниць.

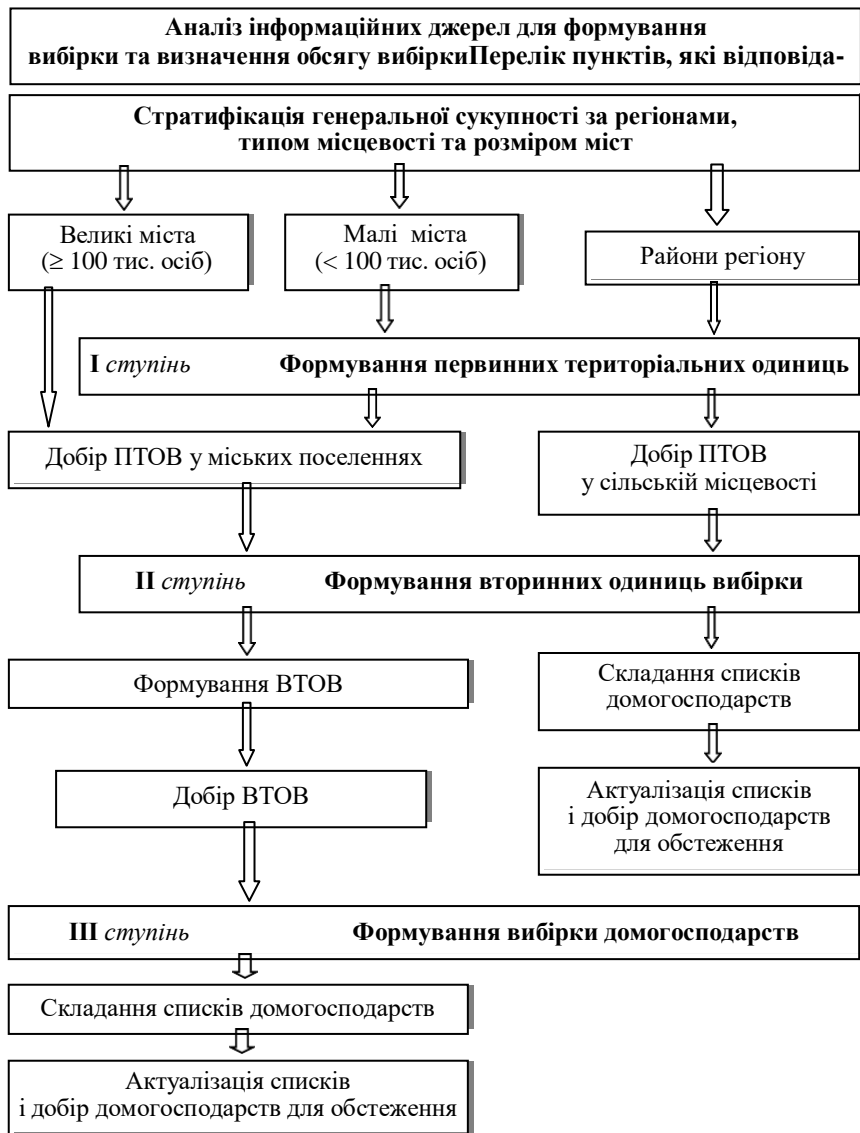


Рис. 8.3. Схема формування вибірки для базових вибіркових обстежень населення (домогосподарств) у 2014–2018 роках

Крок (інтервал) добору визначається діленням загальної чисельності населення на обсяг вибірки, тобто $(430,0 : 7) \cong 61$ тис. осіб. Перший інтервал містить вибіркові одиниці з номерами від 1 до h , другий — від h до $2h$ і т. д. У межах інтервалу за схемою випадкового добору визначаються вибіркові номери: перший номер — множенням випадкового числа rnd на крок добору, наступний — додаванням інтервалу вибірки до першого вибіркового номера і т. д. Одиниця списку, на яку припадає вибірковий номер, включається у вибірку для обстеження.

У табл. 8.8 наведено фрагмент упорядкованого списку міських поселень умовного регіону. Нехай випадкове число $rnd = 0,72$, тоді початковий вибірковий номер буде $(61,0 \cdot 0,72) \cong 44$, наступні номери: $(44 + 61) = 105$; $(105 + 61) = 166$ і т. д. Згідно з цими номерами з наведеного фрагмента списку малих міст регіону до вибірки потрапляють міста: А, С, F, S.

Таблиця 8.8

**ФОРМУВАННЯ ВИБІРКИ ТЕРИТОРІАЛЬНИХ ОДИНИЦЬ У СТРАТІ
«МАЛІ МІСТА» (ЗА ЧИСЕЛЬНІСТЮ НАСЕЛЕННЯ)**

Місто	Кількість одиниць спостереження, тис. осіб		Вибірковий номер	Дібрані одиниці списку
	Локальна	Кумулятивна		
А	48	48	44	А
В	43	91		
С	36	127	105	С
D	34	161		
F	31	192	166	F
H	26	218		
S	17	235	227	S
...
W	2	430		
Разом	430	X	X	X

Другим етапом формування вибірки є добір локальних територіальних одиниць: у містах — це інструкторські дільниці, у сільській місцевості — сільські ради. Упорядкування локальних територіальних одиниць (ВТОВ у містах і ПТОВ у сільській

місцевості) здійснюють за принципом «географічного серпантину», що вможливує рівномірне охоплення території і зменшує витрати на обстеження. Для їх формування необхідно мати картографічний матеріал (карти, карти-схеми, плани), в якому відображаються межі локальних територіальних одиниць.

На рис. 8.4 представлена схема «географічного серпантину», яка накладається на карту території, розміщеної на осі «північ-південь». Лінію серпантину, починаючи з лівого верхнього кута, проводять на схід, після досягнення краю карти має бути поворот на 180° , далі лінія серпантину ведеться зі сходу на захід паралельно попередній лінії. Обов'язковою умовою є відсутність «порожніх проходів», тобто при одному проході лінія має пройти щонайменше через одну територіальну одиницю. За напрямом руху лінії серпантину впорядковуються списки територіальних одиниць.

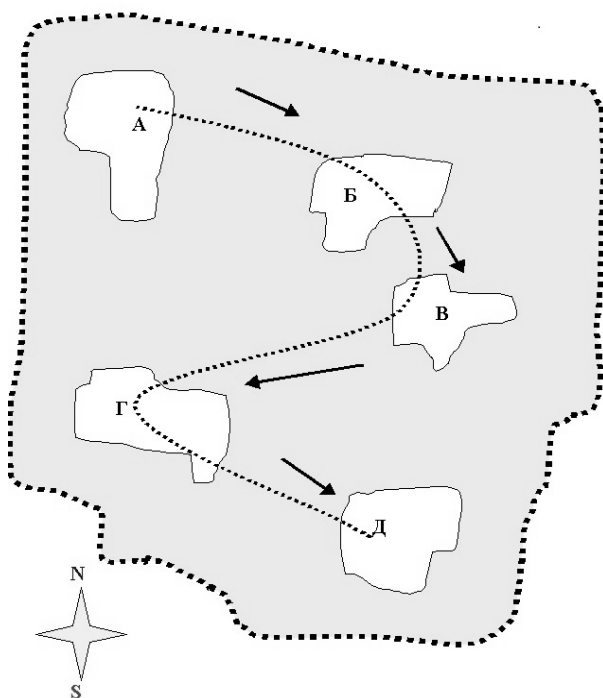


Рис. 8.4. Схема упорядкування списку сільських рад за географічним серпантинном

Подальша процедура добору ВТОВ у страті «малі міста» аналогічна добору несаморепрезентативних ПТОВ. У сільській місцевості розмір ПТОВ має бути не менше встановленого мінімального розміру локальної одиниці добору $n_{min} = 300$ домогосподарств. Якщо розмір локальної одиниці менший мінімального, її об'єднують з іншими локальними одиницями, які географічно знаходяться поруч, допоки не буде забезпечено мінімально необхідний розмір.

Послідовність формування вибірки локальних територіальних одиниць (ВТОВ) подано на рис. 8.5.

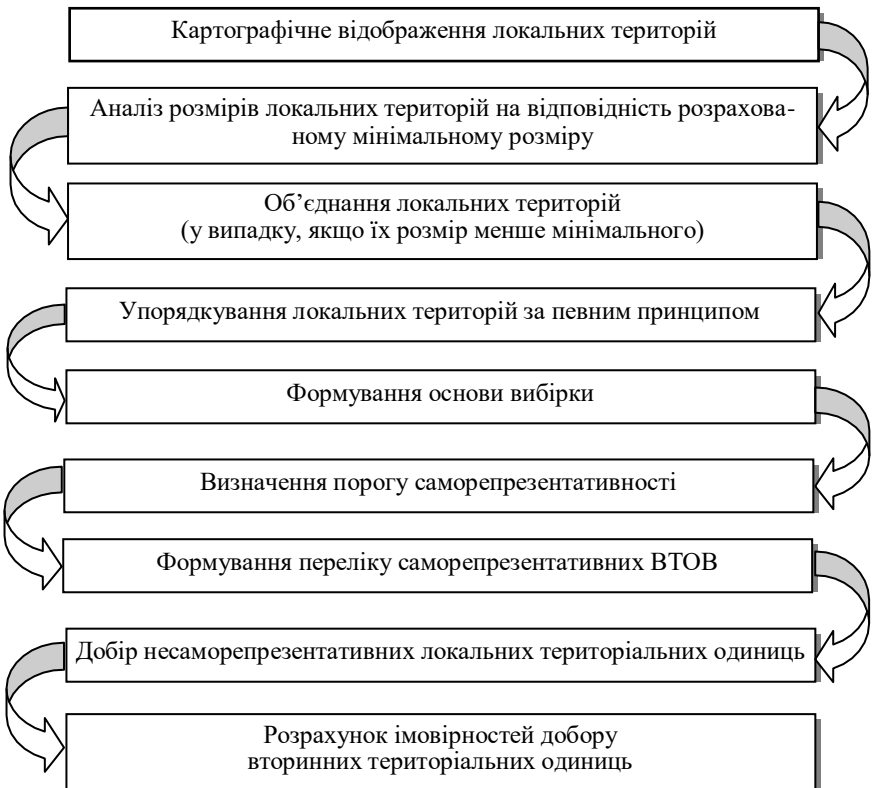


Рис. 8.5. Послідовність формування вибірки локальних територіальних одиниць

На третьому етапі в межах дібраних локальних територій (ВТОВ у містах і ПТОВ у сільських районах) добір домогосподарств для обстеження здійснюють за схемою систематичної вибірки. Як основу вибірки використовують: у містах — списки адрес домогосподарств, які були складені для попереднього добору, у сільській місцевості — дані погосподарських книг. У зв'язку зі змінами, які постійно відбуваються в списках адрес домогосподарств і призводять до певної невідповідності їх реальному стану, обов'язковою процедурою є актуалізація списків і внесення в них необхідних змін. Саме актуалізація списків забезпечує якісне функціонування вибіркової сукупності домогосподарств упродовж усього терміну дії територіальної вибірки.

Упорядкування домогосподарств у міських поселеннях здійснюється поетапно: спершу за абеткою впорядковуються назви вулиць, потім — за порядком зростання номери будинків, у будинках — за порядком зростання номери квартир. У сільській місцевості списки домогосподарств зазвичай складають за тим порядком, як вони записані у книгах погосподарського обліку.

Отже, у багатоступеневій вибірці списки одиниць спостереження (домогосподарств) використовуються лише на останньому ступені. На попередніх ступенях добору вибірковими є територіальні одиниці різного порядку (регіони, міста, сільські райони). Це дає можливість, як і в стратифікованій вибірці, отримати зведені результати обстеження за окремими територіальними одиницями і областями дослідження.

Зауважимо, що при доборі домогосподарств, аби уникнути виходу за межі списку, враховують лише цілу частину вибіркового номера, незалежно від значення дробової частини і правил округлення. Наприклад, при округленні до двох знаків після коми, вибіркового номеру 74,02 чи 74,98, до вибірки включається домогосподарство, яке у списку під номером 74.

На етапі проектуванні вибіркового обстеження визначається надійність, з якою необхідно отримати оцінки основних показників як для національного рівня, так і для рівня окремих адміністративно-територіальних одиниць. Мірою надійності оцінок показників використовують відносну стандартну похибку вибірки V_{μ} . На державному рівні вибіркові оцінки показників вважаються надійними при значеннях похибки не більше 5 %, на регіональному — не більше 10 %.

Вправи та запитання для самоконтролю

1. Поясніть характерні риси і особливості формування багатоступеневої вибірки.

2. Від чого залежить похибка вибірки двоступеневого добору?

3. Поясніть сутність і послідовність формування багатоступеневої вибірки за схемою добору з імовірностями, пропорційними розміру проміжних вибірових одиниць (PPS).

4. Чи можна оцінити точність середньої у багатоступеневій вибірці способом PPS, якщо розміри первинних одиниць характеризуються відносними величинами?

5. Як визначити оптимальне співвідношення обсягів вибірки на різних ступенях добору?

6. Кошторис вибіркового обстеження становить $C = 30\,000$ грн, одиничні витрати на обстеження на першій фазі $c' = 10$ грн, на другій — $c'' = 15$ грн. Коефіцієнт кореляції між ознаками $\delta = 0,88$. Який добір ефективніший — однофазний чи двофазний. За умови організації двофазної вибірки визначте оптимальне розміщення вибірки за фазами.

7. Під час організації вибіркового обстеження якості рибних консервів інспектор спирається на дані попередніх обстежень, за якими загальна дисперсія ознаки якості становить $s^2 = 8$, коефіцієнт внутрішньогніздової кореляції — $0,2$, витрати часу для того, щоб дібрати і відкрити ящик — 2 хв., а для обстеження випадково дібраної банки консервів — $0,5$ хв. Визначте оптимальну кількість ящиків і оптимальну кількість банок, які необхідно дібрати із одного ящика, щоб забезпечити похибку вибірки на рівні $\mu = 0,1$ хв. з мінімальними витратами часу на перевірку.

8. З метою перевірки якості виробів (відповідності стандартам) на комбінаті будматеріалів проведено двоступеневу вибірку партії виробів, які упаковані в ящики по 100 штук у кожному. Витрати часу на обстеження ящика в 4 рази перевищують витрати на обстеження окремих виробів. За даними попередніх обстежень компоненти дисперсії становили: $s_B^2 = 1,5$; $s_W^2 = 6,0$. Визначте, скільки треба дібрати ящиків і скільки в кожному з них перевірити виробів, щоб забезпечити точність оцінювання на рівні стандартної похибки $\mu = 0,3$ з мінімальними витратами часу на перевірку.

9. Телекомунікаційна компанія після встановлення нового обладнання на власних базових станціях регіону (обладнання дозволяє посилювати сигнал рухомого радіозв'язку) планує провести вибірову перевірку зашумленості частотних ліній у зоні безпосередньої дії базових станцій, а також якість роботи передавальних антен. Це передбачає виїзд спеціаліста на дібрані об'єкти і дослідження окремих антен спеціальним приладом. Обстеження здійснюється за схемою двоступеневого добору (станція — антена). Допустимі витрати на обстеження сумарно

становлять 17 000 грн. Витрати на обстеження гнізда (базової станції) становлять 420 грн, обстеження однієї передавальної антени — 130 грн. Коефіцієнт внутрішньогніздової кореляції $r = 0,26$. Визначте оптимальний розмір гнізда і кількість гнізд.

10. Скільки ступенів добору застосовують при формуванні вибіркової сукупності базових обстежень домогосподарств? Обґрунтуйте основу вибірки на кожному ступені добору.

11. Визначте поріг саморепрезентативності малих міст умовного регіону, якщо: обсяг генеральної сукупності $N = 255\,530$ домогосподарств; обсяг вибіркової сукупності $n = 202$ домогосподарств; навантаження статистика $m = 18$ домогосподарств; середній розмір домогосподарства (осіб) $m_{\text{ср}} = 2,45$ особи.

12. Сформуйте територіальну вибірку для обстеження домогосподарств у містах області за принципом «географічного серпантину». Поріг саморепрезентативності — 104,5 тис. осіб. Випадкове число визначте за допомогою генератора випадкових чисел.

Місто	Тис. осіб	Місто	Тис. осіб
Білопілья	60,8	Преслав	29,6
Оленівка	57,3	Вишневе	40,9
Бережани	40,6	Подільськ	243,0
Ізяслав	50,9	Іваношль	36,8
Городище	48,0	Кам'янка	53,1

13. За даними заддання 12 сформуйте територіальну вибірку за розміром (чисельністю населення), використовуючи поріг саморепрезентативності 104,5 тис. осіб і визначене випадкове число.

Порівняйте структуру вибірок територіальних одиниць, сформованих за розміром і за принципом «географічного серпантину», зробіть висновки.

14. Із сукупності обсягом 10 одиниць доберіть для обстеження 4 одиниці з імовірністю, пропорційною їх розміру, випадкове число дорівнює 0,65.

Одиниця сукупності	Відносний розмір одиниці	Одиниця сукупності	Відносний розмір одиниці
1	0,15	6	0,08
2	0,08	7	0,10
3	0,05	8	0,05
4	0,13	9	0,10
5	0,15	10	0,13

15. Поясніть принцип упорядкування й особливості формування основи вибірки локальних територіальних одиниць.

Рекомендована література

1. *Гладун О.М.* Вибіркові обстеження населення: методологія, методика, практика : Монографія. — Ніжин: ТОВ «Видавництво «Аспект — Поліграф», 2008. — 344 с.

2. Методика формування вибірових сукупностей для проведення у 2014–2018 роках вибірових обстежень населення (домогосподарств): умов життя домогосподарств, економічної активності населення та сільськогосподарської діяльності населення у сільській місцевості / Державна служба статистики України. 2013. URL: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2013/37/metod_vyb_suk.zip

3. Методологічні положення щодо формування вибірових сукупностей для проведення органами державної статистики України вибірових обстежень населення (домогосподарств) / Державна служба статистики України. 2018. URL: zakon.rada.gov.ua/rada/file/text/62/f474085n19.pdf

4. Составление планов выборки для обследований домашних хозяйств: практические рекомендации. — ООН, Нью-Йорк, 2010. — 256 с.

РОЗДІЛ 9

ПОВТОРНІ СТАТИСТИЧНІ ОБСТЕЖЕННЯ І ПРИНЦИПИ РОТАЦІЇ

Схеми ротації одиниць вибірки
Оцінювання результатів ротаційної вибірки і забезпе-
чення їх надійності
Особливості панельних обстежень

Після вивчення цього розділу студент має

- *знати*: схеми ротації, які застосовують для підвищення якості результатів обстеження; логіку розрахунку дизайн-ефекту ротаційної вибірки; сутність і сферу використання панельних обстежень;
- *уміти*: обирати схему ротації; визначати обсяги ротаційної вибірки з урахуванням вимог щодо рівня надійності цільових показників, розподіляти їх за стратами; отримувати надійні вибіркові оцінки середніх значень цільових показників та зміни середніх від одного раунду обстеження до іншого.

Схеми ротації одиниць вибірки

У зв'язку з необхідністю моніторингу швидкоплинних соціально-економічних явищ і процесів поширюється практика повторних вибіркових обстежень. Для підвищення якості результатів вибіркових обстежень, які проводяться на постійній основі, застосовують спеціальну процедуру — ротацію одиниць вибірки. Сутність ротації полягає в заміні частини діючої вибіркової сукупності за певною схемою в кожному черговому раунді вибіркового обстеження. Задля здійснення цієї процедури вибіркова сукупність розбивається на окремі підвибірки — ротаційні групи, алгоритм зміни яких є складовою дизайну вибірки, регламентується методологією кожного окремого обстеження і обов'язково визначається в технічному завданні на формування вибірки.

Застосування процедури ротації вибіркових одиниць, з одного боку, забезпечує оновлення вибіркової сукупності для кожного раунду обстеження, а з іншого — зберігає незмінною певну її частину, що дає можливість надійніше виміряти зміни оцінок пока-

зників у динаміці, не перевантажуючи будь-яку окрему групу одиниць тривалим періодом обстеження.

Вибір конкретної схеми ротації залежить від мети обстеження, можливості організаційно забезпечити підтримку ротаційної схеми, вимог до надійності вибіркового оцінок показників тощо. Як приклад розглянемо *Методику формування ротаційної вибірки для проведення поточних обстежень підприємств із питань статистики праці* [Методика формування вибірки]. Головна мета обстеження — отримати надійну інформацію щодо оплати праці та інших аспектів соціально-трудових відносин. Одиницею спостереження є місцева одиниця (підприємство або його частина), розташована в географічно визначеному місці. Генеральну сукупність обстеження становить ідентифіковані в ЄДРПОУ звітні одиниці із середньою кількістю найманих працівників від 10 до 49 осіб включно. Вибіркова сукупність формується з використанням процедури стратифікованого випадкового добору. Стратифікація здійснюється за основним видом економічної діяльності (згідно з КВЕД) і за територіальною належністю звітної одиниці (відповідно до адміністративно-територіального устрою країни).

Термін дії вибірки становить 7 років. З метою кращого охоплення обстеженням генеральної сукупності та зниження звітного навантаження на респондентів застосовують таку схему ротації одиниць вибірки, за якою у кожній страті звітних одиниць добирається сім ротаційних груп, кожна з них представлена двома рівними за розміром частинами. Щороку після актуалізації основи вибірки одна частина ротаційної групи (50 % вибіркової сукупності) заміщується такою самою частиною нової ротаційної групи. Тобто щороку в обстеженні беруть участь дві ротаційні групи, а дібрані до певної ротаційної групи звітні одиниці, за винятком першої, обстежуються два роки поспіль із подальшими п'ятирічними «канікулами» (рис. 9.1).

Загальний обсяг вибірки n визначається з урахуванням відповідного дизайн-ефекту за формулою:

$$n = deff \cdot s^2 \left(\frac{100}{\bar{y} \cdot \bar{y}} \right)^2,$$

де \bar{y} — середній рівень заробітної плати в генеральній сукупності;

s^2 — дисперсія заробітної плати в генеральній сукупності;

V_{μ} — відносна стандартна похибка вибіркової оцінки середньої заробітної плати, яка відповідає встановленим вимогам щодо надійності результатів обстеження на національному рівні;

$deff$ — значення дизайн-ефекту при оцінюванні середніх доходів та витрат домогосподарств та економічної активності населення за рекомендаціями міжнародних експертів приймаються на рівні 2,5.

		Ротаційна група						
		1	2	3	4	5	6	7
Рік	2010	■	■					
	2011		■	■				
	2012			■	■			
	2013				■	■		
	2014					■	■	
	2015						■	■
	2016	■						■
	2017		■					
	2018			■				
	2019				■			
	2020					■	■	
	2021						■	■
	2022							■

Рис. 9.1. Схема ротації одиниць вибірки в обстеженні підприємств з питань статистики праці

Середній рівень і дисперсія заробітної плати (\bar{y} , s^2) визначаються не для всієї сукупності малих підприємств (10–49 працівників), а після вилучення 0,5 % найменших і 0,5 % найбільших значень упорядкованого за зростанням ряду середньої заробітної плати. Процедура обрізання (trimming) крайніх значень упорядкованого ряду сприяє зменшенню впливу на оцінки результатів вибірки нетипових за цільовими ознаками одиниць.

Необхідні для надійного оцінювання заробітної плати обсяги вибірки розраховують для кожного j -го регіону — n_j і для кожного i -го виду економічної діяльності (на національному рівні)

— n_i . По Україні обсяги вибірки n^Σ визначають як суму отриманих обсягів:

а) для регіонів $n_j^\Sigma = \sum_j n_j$, де j — код регіону за КОАТУУ;

б) для видів діяльності $n_i^\Sigma = \sum_j n_{ij}$, де i — код виду економічної діяльності за КВЕД (секції, підсекції).

Більший із розрахованих обсягів, отриманих для кожної групи звітних одиниць, після коригування визнається як теоретичний $n^\Sigma = \max(n_j; n_i)$ і розподіляється за стратами (регіонами та видами економічної діяльності) з урахуванням кількості звітних одиниць в генеральній сукупності та дисперсії середньої заробітної плати у кожній страті.

Розміщення вибірки	
за регіонами	за секціями КВЕД
$n_j^* = n^\Sigma \cdot \frac{\sqrt[3]{N_j \cdot s_j}}{\sum_j \sqrt[3]{N_j \cdot s_j}}$	$n_j^* = n^\Sigma \cdot \frac{\sqrt[3]{N_i \cdot s_i}}{\sum_i \sqrt[3]{N_i \cdot s_i}}$
<p>де N_j та s_j — загальна кількість звітних одиниць та стандартне відхилення заробітної плати працівників за звітними одиницями j-го регіону; N_i та s_i — загальна кількість звітних одиниць та стандартне відхилення заробітної плати працівників за звітними одиницями i-го виду економічної діяльності (на національному рівні).</p>	

Слід зазначити, що розміщення вибірки за наведеними формулами є проміжним варіантом між відомим способом оптимального розміщення Неймана та розміщенням за стандартним алгоритмом. Обраний спосіб розміщення вибірки забезпечує більш надійне оцінювання показників у межах окремих видів діяльності, зменшує ймовірність реалізації випадків, коли обсяг вибірки в страті може перевищувати обсяг генеральної сукупності загалом.

Передбачається, що за результатами обстеження підприємств із питань статистики праці рівень надійності оцінювання основних показників буде відповідати таким критеріям [Методика формування вибірки, 2006]:

- на національному рівні відносна стандартна похибка оцінок показників не перевищуватиме 5%;
- за сукупністю звітних одиниць у кожному з регіонів та за кожним із видів економічної діяльності — 10 %;

– за сукупністю звітних одиниць за видами економічної діяльності в межах регіонів та адміністративно-територіальних одиниць — 15 %.

Зазначені критерії надійності статистичної інформації відповідають рекомендаціям експертів та міжнародній практиці.

Схеми ротації вибірових одиниць є обов'язковим елементом дизайну базових вибірових обстежень населення (домогосподарств), які проводяться на постійній основі. Ротація забезпечує підвищення якості результатів обстеження населення за рахунок зниження навантаження на респондентів, оскільки останні перебувають в обстеженні визначений схемами ротації час (певну кількість періодів обстеження). Крім того, як уже зазначалося, ротація одиниць вибірки забезпечує надійніше вимірювання ступеню зміни оцінок показників. Специфіку ротаційної схеми в базових обстеженнях населення розглянемо на прикладі обстеження економічної активності населення (ОЕАН) у циклі 2014–2018 років. Цільовими показниками, надійність оцінювання яких необхідно забезпечити, є: а) рівень безробіття та б) рівень зайнятості населення віком 15–70 років, %.

Згідно з методологією обстеження економічної активності населення ротація здійснюється за схемою 3 — 9 — 3, тобто кожне дібане домогосподарство обстежується три місяці поспіль, потім перерва у дев'ять місяців, після чого обстежується ще три місяці. Таким чином, кожного місяця фахівець з інтерв'ювання проводить обстеження у шести ротаційних групах, з них дві групи щомісяця замінюються на нові. Наприклад, у першому кварталі першого року впровадження схеми ротації в обстеженні беруть участь групи з такими номерами:

січень — 1, 2, 3, 13, 14, 15;

лютий — 2, 3, 4, 14, 15, 16;

березень — 3, 4, 5, 15, 16, 17.

З урахуванням схеми ротації визначають теоретичні обсяги вибіркової сукупності, яку необхідно сформувати на рік, а також мінімальний розмір територіальної одиниці. За регіонами теоретичні обсяги вибірки n_j^* визначаються з урахуванням вимог щодо рівня надійності оцінюваного показника на регіональному рівні ($V_{\mu, st} = 10\%$):

$$n_j^* = \sum_j \left(n_j \cdot \frac{v^2}{v_{\mu, st}^2} \right),$$

де n_j — фактичний обсяг вибірки для j -го регіону за результатами попереднього обстеження;

V_{μ} — середнє фактичне значення відносної стандартної похибки за результатами попереднього обстеження;

$V_{\mu, st}$ — рівень надійності цільового показника за вимогами до формування вибірки.

Технічним завданням встановлені рівні надійності цільових показників з урахуванням дизайн-ефектів: на національному рівні — 3 %, на регіональному — 10 %. Загальний теоретичний обсяг вибірки (кількості домогосподарств) визначається як сума теоретичних обсягів вибірки за регіонами. На рис. 9.2 схематично представлено порядок визначення теоретичних обсягів вибірки за регіонами і розподілу їх за stratami в обстеженні економічної активності населення.

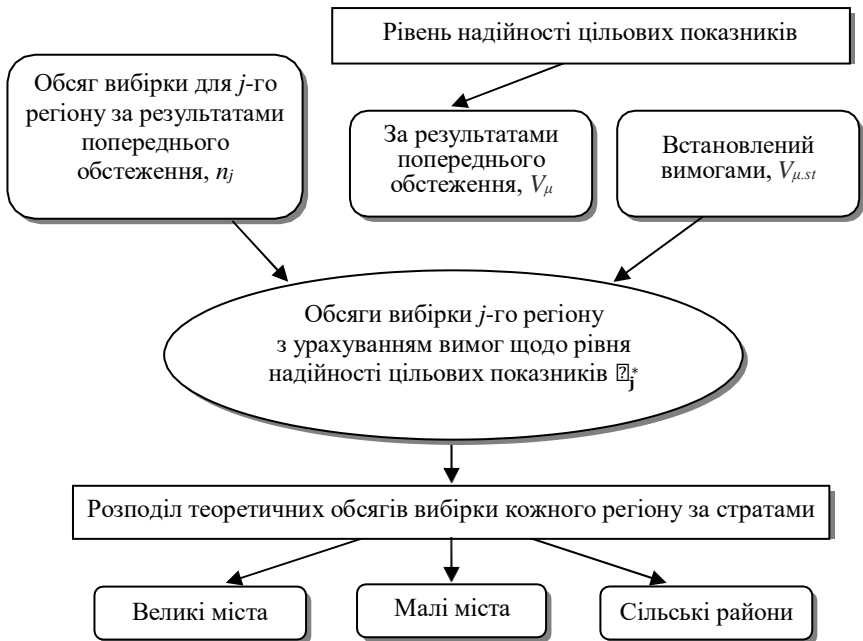


Рис. 9.2. Загальний порядок визначення теоретичних обсягів вибірки за регіонами і розподілу їх за stratami в обстеженні економічної активності населення

Розподіл теоретичних обсягів вибірки кожного регіону за стратами: «великі міста», «малі міста», та «сільські райони» (останні розглядаються як одна страта) здійснюють пропорційно чисельності постійного населення регіону з урахуванням надійності оцінювання показників за формулою:

$$n_{w,j} = n \cdot \frac{p_{w,j} \cdot s_{w,j}}{\sum_j p_{w,j} \cdot s_{w,j}},$$

де $n_{w,j}$ — обсяг вибірки для w -ої страти j -го регіону ($w = 1, 2, 3$);

n_j — обсяг вибірки для j -го регіону;

$p_{w,j}$ — частка загальної чисельності постійного населення, що припадає на w -ту страту j -го регіону;

$s_{w,j}$ — оцінка стандартного відхилення цільового показника w -ої страти j -го регіону.

Порядок розрахунку обсягів вибірки за стратами розглянемо на прикладі регіону, скоригований теоретичний обсяг вибірки якого становить 370 домогосподарств. Структуру постійного населення за стратами й оцінки стандартних відхилень наведено в табл. 9.1.

Таблиця 9.1

ДО РОЗРАХУНКУ ОБСЯГІВ ВИБІРКИ ЗА СТРАТАМИ В J-МУ РЕГІОНІ

Страта	Частка постійного населення, $p_{w,j}$	Стандартне відхилення, $s_{w,j}$	$p_{w,j} s_{w,j}$	Обсяг вибірки, $n_{w,j}$
Великі міста	0,267	2106,5	562,43	115
Малі міста	0,360	1487,3	535,43	110
Сільські райони	0,373	1898,5	708,14	145
Разом	1,000	X	1806,10	370

Згідно з даними:

$$n_1 = 370 \frac{562,43}{1806,10} \approx 115 \text{ дг}$$

$$n_2 = 370 \frac{535,43}{1806,10} = 110 \text{ дг}$$

$$n_3 = \frac{708,14}{1806,10} \equiv 145 \text{ дг}$$

Якщо у вимогах до формування вибірки вказано очікуваний коефіцієнт відповідей $k_{гр}$, мінімальний розмір територіальної одиниці коригується:

$$n'_{\min} = \frac{n_{\min}}{k_{гр}}$$

Скоригований теоретичний обсяг вибірки за типами населених пунктів по країні в цілому розраховується як сума скоригованих теоретичних обсягів вибірки за всіма регіонами.

Для розміщення скоригованого обсягу вибірки за втратами необхідно врахувати, що відповідно до схеми ротації, вибірка складається із двох частин, одна з яких щороку замінюється на нову; при цьому на весь семирічний цикл обстеження у кожній страті необхідно дібрати сім ротаційних груп.

У випадку проведення в одній територіальній одиниці кількох вибіркових обстежень мінімальний розмір територіальної одиниці визначається сумою мінімальних розмірів, розрахованих для кожного з цих обстежень незалежно від можливого спільного використання для обстежень окремих частин вибірок.

Оцінювання результатів ротаційної вибірки і забезпечення їх надійності

Схеми ротації в обстеженнях соціально-економічних явищ різняться, із кожним новим раундом обстеження може змінюватися дизайн вибірки, проте концептуально будь-яка схема ротації дає можливість:

- отримувати надійні вибіркові оцінки середніх значень цільових показників;
- оцінювати зміни середніх від одного раунду обстеження до іншого;
- досліджувати тенденції показників і здійснювати прогнозування.

Метою більшості вибіркових обстежень є поточне оцінювання параметрів сукупності, передусім це стосується середніх рівнів тих показників, які швидко змінюються. У тих випадках, коли необхідно виявити вплив на динаміку цільових показників пев-

них факторів, скажімо, інфляції чи законодавчих актів, актуальнішим є оцінювання зміни середніх.

Практично забезпечити однакову структуру вибірки при переході від одного раунду обстеження до іншого неможливо. Склад сукупності змінюється під впливом різних причин, пов'язаних з демографічними процесами, із соціальною мобільністю респондентів, з відмовами брати участь в обстеженні тощо. Структурні зрушення у сукупності респондентів спричиняють появу систематичних похибок і зсувнення вибірових оцінок.

При оцінюванні поточних значень середніх \bar{x} за даними послідовних вибірок без заміщення і при повному заміщенні припускають, що дисперсія s_x^2 не змінюється, а отже, похибка вибірки визначається так, як для простого випадкового добору. Складнішою є процедура оцінювання при другому, проміжному варіанті послідовної вибірки, для якого характерне часткове заміщення вибірових одиниць при переході від одного раунду обстеження до іншого.

Розглянемо методологічні принципи формування ротаційної вибірки обсягом n для двох послідовних моментів. При формуванні кожної наступної вибірки в ній залишається m одиниць попередньої, а q одиниць заміщуються новими. Декомпозицію середніх рівнів послідовних вибірок схематично представлено на рис. 9.3.

1-а вибірка	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}		
	x_1'		x_1''							
2-а вибірка			x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}
			\bar{x}_2'						\bar{x}_2''	
Загальні середні			\bar{x}_1							
			\bar{x}_2							

Рис. 9.3. Схема декомпозиції середніх величин ротаційної вибірки

Основні характеристики послідовних вибірок:

Обсяг вибірок	$n_1 = n_2 = 8$
Збіжні одиниці вибірок	$m = 6$
Незбіжні одиниці (заміщення)	$q = 4$
Частка збіжної частини вибірових сукупностей	$d = m : n = 6 : 8 = 0,75$
Частка заміщення (ротації)	$1 - d = 1 - 0,75 = 0,25$
Середні значення ознаки першої вибірки: загалом	\bar{x}_1
для частини збіжних одиниць	\bar{x}_1''
для частини заміщених одиниць	\bar{x}_1'
Середні значення ознаки другої вибірки: загалом	\bar{x}_2
для частини збіжних одиниць	\bar{x}_2'
для частини нових одиниць	\bar{x}_2''

Збіжні і незбіжні частини другої вибірки дають незалежні оцінки середньої \bar{x}_2' та \bar{x}_2'' . Оцінка середньої збіжної частини визначається методом регресії для двофазного добору. Допоміжною інформацією слугують дані попередньої вибірки. Формули оцінок середніх і дисперсії поточної вибірки наведено в табл. 9.2.

Таблиця 9.2

ОЦІНКИ СЕРЕДНІХ І ДИСПЕРСІЙ ПО ЗБІЖНІЙ І НЕЗБІЖНІЙ ЧАСТИНАХ РОТАЦІЙНОЇ ВИБІРКИ

Частина вибірки	Оцінка середньої	Дисперсія
Незбіжна	\bar{x}_2''	$\frac{s^2}{q} = \frac{1}{W_q}$
Збіжна	$\bar{x}_2' + b(\bar{x}_1 - \bar{x}_1')$	$\frac{s^2(1-r^2)}{m} + r^2 \frac{s^2}{n} = \frac{1}{W_m}$

Оцінку поточного значення середньої \bar{x}_2 з урахуванням ротації сукупності визначають як композиційну оцінку середніх по

збіжній і незбіжній частинах другої вибірки з вагами, оберненими до їхніх дисперсій:

$$\bar{x}_2 = \phi \bar{x}'_2 + (1 - \phi) \bar{x}''_2,$$

де ϕ і $(1 - \phi)$ — ваги незалежних оцінок середніх.

За умови мінімізації похибки оцінок вага ϕ визначається за формулою:

$$\phi = \frac{W_q}{W_q + W_m}.$$

Похибка вибірки для середньої \bar{x}_2 становить:

$$\mu_y^2 = \frac{1}{W_q + W_m} = \frac{s^2(n - qr^2)}{n^2 - q^2r^2}.$$

Ця формула підтверджує, що при $q = 0$ (вибірки повністю збігаються) і при $q = n$ (повне заміщення) похибка вибірки для середньої має одне й те саме значення $\frac{S^2}{n}$. Мінімізуючи похибку вибірки по q , можна визначити оптимальне значення частки заміщення:

$$\frac{q}{n} = \frac{1}{1 + \sqrt{1 - r^2}}; \quad \frac{m}{n} = \frac{\sqrt{1 - r^2}}{1 + \sqrt{1 - r^2}}.$$

Якщо підставимо оптимальне значення q у формулу похибки вибірки, дістанемо мінімальне значення похибки:

$$\mu_y^2 = \frac{s^2}{2n} \left[1 + \sqrt{1 - r^2} \right].$$

У табл. 9.3 [Кокрен, 1976: с. 368] наведено для різних значень r оптимальний процент збіжних одиниць та відносний виграш у точності порівняно з вибіркою повного заміщення.

Оптимальна частка збіжних одиниць в ротаційній вибірці не перевищує 46% (при $r = 0,5$) і постійно зменшується зі зростанням коефіцієнта кореляції r . Виграш у точності досить помірний і практично не залежить від частки заміщення вибіркової сукупності. Обидва варіанти заміщення ($d = 1/3$ і $d = 1/4$) дають приблизно однаковий результат.

Таблиця 9.3

ОПТИМАЛЬНИЙ ПРОЦЕНТ ЗБІЖНИХ ОДИНИЦЬ РОТАЦІЙНОЇ ВИБІРКИ

r	Оптимальний процент збіжних одиниць	Виграш у точності (%) при	
		$\frac{m}{n} = \frac{1}{3}$	$\frac{m}{n} = \frac{1}{4}$
0,5	46	7	6
0,6	44	11	9
0,7	42	17	15
0,8	38	25	23
0,9	30	39	39
0,95	24	50	52
1,0	0	67	75

Як зазначалося у 9.1, в ротаційній вибірці зі ступенем заміщення $(1-d)$ оцінки середніх послідовних вибірок розраховують так:

■ для першої вибірки як зважена з незалежних оцінок збіжної і незбіжної частин:

$$\bar{x}_1 = d\bar{x}'_1 + (1-d)\bar{x}''_1,$$

■ для другої вибірки — методом регресії для двофазного добору:

$$\bar{x}_2 = \bar{x}'_2 + b(\bar{x}_1 - \bar{x}'_1) = \bar{x}'_2 + b(1-d)(\bar{x}''_1 - \bar{x}'_1),$$

де коефіцієнт регресії b визначається за даними збіжної частини вибірки.

На підставі наведених формул оцінку зміни середньої можна подати так:

$$\bar{x}_2 - \bar{x}_1 = \bar{x}'_2 - \bar{x}'_1 - (1-d)(1-b)(\bar{x}''_1 - \bar{x}'_1).$$

У практиці вибірових обстежень при оцінюванні зміни середніх частіше використовують коефіцієнт кореляції r , який порівняно з коефіцієнтом регресії більш сталий і при оцінюванні має меншу похибку. Водночас модифікується алгоритм розрахунку оцінки зміни середніх. Він має вигляд середньозваженої з відхилень оцінок середніх $(\bar{x}'_2 - \bar{x}'_1)$ і $(\bar{x}''_2 - \bar{x}''_1)$. Вагами виступають,

відповідно, $\frac{d}{1-(1-d)r}$ і $\frac{(1-d)(1-r)}{1-(1-d)r}$, тобто:

$$\bar{x}_2 - \bar{x}_1 = \frac{d}{1-(1-d)r} (\bar{x}'_2 - \bar{x}'_1) + \frac{(1-d)(1-r)}{1-(1-d)r} (\bar{x}''_2 - \bar{x}''_1).$$

Наприклад, визначимо оцінку зміни середньої за наведеними у табл. 9.4 результатами послідовних вибірових обстежень 12 страхових компаній щодо суми виплат страхового відшкодування.

Таблиця 9.4

СУМИ ВИПЛАТ СТРАХОВОГО ВІДШКОДУВАННЯ, ТИС. ГРН

Страхова компанія	Листопад	Грудень	Страхова компанія	Листопад	Грудень
1	8,8		7	9,1	9,2
2	8,9		8	9,4	9,3
3	9,3		9	9,3	9,5
4	9,2	9,3	10		9,2
5	9,3	9,4	11		9,3
6	8,9	9,1	12		9,1

Усі необхідні розрахунки для оцінки зміни середньої суми страхового відшкодування за грудень порівняно з листопадом наведено в табл. 9.5.

Таблиця 9.5

ДО РОЗРАХУНКУ ОЦІНКИ ЗМІНИ СЕРЕДНІХ ЗА РОТАЦІЙНОЮ ВИБІРКОЮ

Листопад (перша вибірка)	Грудень (друга вибірка)	Відхилення середніх
$\sum x_1'' = 55,2$	$\sum x_2'' = 55,8$	X
$\sum x_1' = 27,0$	$\sum x_2' = 27,6$	X
$\bar{x}_1'' = 9,2$	$\bar{x}_2'' = 9,3$	$9,3 - 9,2 = 0,1$
$\bar{x}_1' = 9,0$	$\bar{x}_2' = 9,2$	$9,2 - 9,0 = 0,2$
$\bar{x}_1 = 9,13$	$\bar{x}_2 = 9,27$	$9,27 - 9,13 = 0,14$
Частка збіжної частини вибірки		$d = 6 : 9 = 0,67$
Частка незбіжної частини вибірки (заміщення)		$1 - d = 3 : 9 = 0,33$

Необхідні суми для розрахунку лінійного коефіцієнта кореляції становлять:

$$\sum (x_1' - \bar{x}_1')^2 = 0,16; \sum (x_1' - \bar{x}_1')(x_2' - \bar{x}_2') = 0,10; \sum (x_2' - \bar{x}_2')^2 = 0,10.$$

Звідси:

$$r = \frac{0,10}{\sqrt{0,16 \cdot 0,10}} = 0,79.$$

Використовуючи наведені у табл. 9.4 частки збіжних і незбіжних частин вибірових сукупностей ($d = 0,67$; $1 - d = 1 - 0,67 = 0,33$), визначимо ваги для оцінки зміни середньої:

$$\frac{d}{1 - (1 - d)r} = \frac{0,67}{1 - 0,33 \times 0,79} = 0,906 ;$$

$$\frac{(1 - d)(1 - r)}{1 - (1 - d)r} = \frac{0,33(1 - 0,79)}{1 - 0,33 \times 0,79} = 0,094 .$$

За таких ваг оцінка зміни середньої дорівнює:

$$\bar{x}_2 - \bar{x}_1 = 0,906 \times 0,10 + 0,094 \times 0,2 = 0,109.$$

Якщо число моментів обстеження 3 і більше, то з'являється можливість застосувати інші варіанти заміщення. Скажімо, в t -ий момент можна отримати вибірку, частина якої збігається з вибіркою в момент $t - 1$, частина — з вибіркою в момент $t - 2$ і т. д. Щоб отримати поточну оцінку, можна застосувати множинну регресію, спираючись на всі вибірки, що збігаються з попередніми моментами. У табл. 9.6 [Кокрен, 1976: с. 371] наведено значення оптимального проценту збіжних одиниць для 2–5 моментів обстеження.

Таблиця 9.6

ОПТИМАЛЬНА ЧАСТКА ЗБІЖНИХ ОДИНИЦЬ РОТАЦІЙНОЇ ВИБІРКИ

t	Процент збіжних одиниць 100 m_t / n при r			
	0,7	0,8	0,9	0,95
2	42	38	30	24
3	49	47	42	36
4	50	49	47	43
5	50	50	49	46
∞	50	50	50	50

Як свідчать дані таблиці, чим щільніший зв'язок, тим менша оптимальна частка збіжних одиниць, а отже, більша частка заміщення. Водночас для всіх наведених значень r оптимальна частка

заміщення зменшується з кожним новим моментом обстеження, досягаючи рівня 50 %.

На практиці передбачається, що витрати на кожне наступне обстеження будуть меншими. Якщо мета дослідження — оцінити зміни середнього чи сумарних значень, доцільно зберігати від одного моменту до іншого близько половини одиниць. Окрім того, слід зберігати частки збіжних одиниць постійними, а не змінювати їх щоразу.

Результати досліджень показують, що збільшення частки збіжних одиниць з $\frac{1}{2}$ до $\frac{3}{4}$ дає невелике зростання похибки вибірки поточної оцінки середньої і досить істотне зменшення похибки оцінки змін. Наприклад, при $r = 0,8$ похибка поточної середньої збільшується приблизно на 5 %, а похибка оцінки зміни середньої зменшується більш як на 20 %.

Отже, якщо необхідно отримати і поточні оцінки, і оцінки змін, прийнятною тактикою заміщення буде збереження в наступний момент обстеження $\frac{2}{3}$, $\frac{3}{4}$ або $\frac{4}{5}$ обсягу попередньої вибірки. У міжнародній практиці обстеження домогосподарств з метою підвищення якості результатів здійснюється щоквартальна ротація респондентів.

Особливості панельних обстежень

Одним з різновидів повторних вибірових обстежень є панельні обстеження. Загальний принцип їх організації полягає в регулярних обстеженнях сформованої на постійній основі вибіркової сукупності з використанням незмінної методики спостереження, що забезпечує порівнянність інформації в часі в тим самим дає змогу можливість відстежувати зміни щодо предмета дослідження від одного раунду обстеження до іншого.

Залежно від мети обстеження і виду панелі одиницями її сукупності можуть бути: а) об'єкти торговельної мережі; б) домашні господарства; в) особи (респонденти), громадська думка і споживчі орієнтації яких аналізуються; г) підприємства, обрані для моніторингу певних процесів; д) групи фахівців у тій чи іншій галузі знань або у сфері бізнесу.

Панель як вибірка сукупність формується методом випадкового або систематичного добору. Аби забезпечити представництво у вибірці всіх типів респондентів, генеральну сукупність

стратифікують за найважливішими її ознаками. Скажімо, при моніторингу соціальних процесів у панелі мають бути пропорційно представлені відповідні демографічні та соціально-економічні типи населення.

Методологія формування панелей передбачає також процес ротації, тобто включення нових учасників панелі замість вибулих. У разі, коли в структурі генеральної сукупності відбулися зміни, задля забезпечення репрезентативності результатів спостереження виникає необхідність коригування структури панелі.

Найчастіше панельні обстеження використовують у маркетингових дослідженнях, у моніторингу ділової активності підприємств, у дослідженнях медіа-аудиторії. Здійснюють їх переважно на комерційній основі спеціалізовані компанії, дотримуючись міжнародних стандартів проведення маркетингових і соціологічних досліджень ICC/ESOMAR⁴.

За предметом дослідження вирізняють п'ять основних типів панелей: торговельні, споживчі, моніторингові, медіапанелі і панелі фахівців-експертів з досліджуваних проблем (табл. 9.6).

Таблиця 9.6

ТИПИ ДОСЛІДНИЦЬКИХ ПАНЕЛЕЙ

Тип панелі	Одиниці сукупності	Цілі формування панелі
Панель об'єктів роздрібно́ї торгівлі	Торговельні підприємства та індивідуальні продавці	Отримувати безперервну інформацію про ринкову ситуацію, ефективність реклами, ціни
Споживча панель домогосподарств	Домогосподарства/ сім'ї або окремі особи	Систематично або періодично отримувати інформацію про попит, поведінку і переваги споживачів
Спеціалізована моніторингова панель	Суб'єкти бізнесу у сфері виробництва товарів та надання послуг	Моніторинг ситуації на спеціалізованих ринках товарів і послуг, цінової політики та систем дистрибуції
Медіапанель	Домогосподарства, окремі особи	Вимірювання цільової теле- та інтернетаудиторій, інших засобів масової інформації
Панель фахівців	Групи фахівців / експертів	Експертні оцінки ситуацій і прогнози розвитку процесу

⁴ Міжнародний кодекс дослідницької етики (ICC / ESOMAR) розроблений Європейською асоціацією дослідників громадської думки і маркетингу (*European Society of Marketing Research Professionals*, ESOMAR) у співпраці з Міжнародною торговою палатою.

За методами дослідження панелі поділяють на традиційні, ґрунтовані на повторних вимірах одних і тих самих параметрів, і на омнібус-панелі, об'єкти дослідження яких постійні в часі, а завдання і програма дослідження змінюються залежно від інформаційних потреб замовників.

Терміни експлуатації панелей варіюють від місяця до кількох років, за цією ознакою їх об'єднують у дві групи: короткострокові (не більше року) і довгострокові (до п'яти років). Панелі різняться також за обсягом вибірки і способами одержання інформації (інтерв'ю, опитувальні листи, анкети, щоденники, спеціальні пристрої). Доцільність використання того чи іншого виду панелі визначається характером вирішуваних завдань і кошторисом. Розглянемо спільні риси та особливості окремих видів панелей.

Торговельна панель (Trading panel) — це панель об'єктів мережі роздрібної торгівлі, предметом дослідження якої є продажі, ціни, запаси певних видів товарів. Мета панельного дослідження — моніторинг ринкової пропозиції як у межах національного ринку, так і в регіональному розрізі. Панель формують як постійну вибірку підприємств роздрібної торгівлі з урахуванням їхнього розміру (обсягу товарообігу) та специфіки асортиментного профілю. Відповідно до структури роздрібної мережі у вибірку включають певну кількість універсамів (супермаркетів), універмагів, спеціалізованих магазинів тощо. На включених до панелі торговельних одиницях регулярно (найчастіше раз на два місяці) фіксуються обсяги продажів товарів певного асортименту, наявні запаси їх на складах і в торговельних залах, ціни і марки товарів, види упаковки тощо.

За результатами панельного обстеження визначають низку показників, необхідних для оперативного коригування асортиментної і цінової політики, виявлення недоліків у товарній пропозиції, пошуку нових маркетингових можливостей, зокрема: місткість і тенденції ринку для кожного сорту і торгової марки; обсяги продажу певного товару, його розподіл за типами магазинів і регіонами; використання тих чи інших способів залучення покупців (знижки, купони, премії); специфіку і розміри упаковки проданих товарів; цінову і бізнес-політику, інші.

Формування й підтримка торговельних панелей вимагають значних витрат, тому цим займаються переважно спеціалізовані фірми, серед яких найвідомішою є дослідницька компанія Nielsen. Це один зі світових лідерів у галузі маркетингових досліджень ринків споживчих товарів повсякденного попиту (fast

moving consumer goods, FMCG). В Україні функціонує філія цієї компанії — Nielsen Ukraine.

Споживча панель (Consumer Scan) — це панель домашніх господарств, які щоденно надають інформацію щодо своїх покупок. На відміну від торговельної панелі споживча дає відповідь на питання про те, хто є покупцем того чи іншого товару.

Формування споживчої панелі передбачає дотримання головних принципів організації вибіркового обстеження, передусім репрезентативності вибірки та регулярності проведення вимірів. В Україні наймасштабніші панельні дослідження домогосподарств проводять компанії GfK Ukraine та Nielsen Holdings. Національна споживча панель формується з урахуванням географічних і демографічних критеріїв. Вибірка домогосподарств панелі базується на інформації органів державної статистики, вона репрезентативна за статтю та віком, за типом поселення (міське і сільське населення), враховує розмір населеного пункту. Дослідження покриває понад 60 груп товарів повсякденного попиту (FMCG categories), зокрема продукти харчування, напої, чистячі/мийні засоби та продукти персональної гігієни.

Кожне дібране до панелі домогосподарство веде щоденник, де фіксуються покупки товарів, передбачених програмою дослідження. Стосовно кожної покупки вказується: марка, виробник, кількість, вага, тип упакування, ціна, дата і місце покупки, характеристики покупця. Отримані за результатами панельного обстеження дані дозволяють робити висновки про обсяги споживання, про зміни купівельної поведінки домогосподарств щодо досліджуваних товарів, про тенденції купівельного попиту, цінові переваги, вподобання і споживчі настрої покупців.

Через інтенсивну зміну ситуації на споживчих ринках сфера застосування Consumer Scan у різних країнах світу розширюється, постійно з'являються різноманітні модифікації панелей, вдосконалюється методика їх формування й інструментарій спостереження. Новітні дослідницькі рішення компаній, які працюють на споживчих ринках, спрямовані на вимірювання купівельної поведінки домогосподарств і прогнозування споживчих настроїв. Провідним показником тенденцій на споживчому ринку слугує індекс споживчих настроїв (Consumer Confidence Index, CCI). В усьому світі його використовують як індикатор потенційної зміни споживчого попиту в наступних періодах. Розрахунок CCI ґрунтується на порівнянні часток позитивних d_p і негативних d_n

відповідей респондентів споживчої панелі: $CCI = (d_p - d_n) + 100$. Значення індексу змінюється в інтервалі від 0 до 200. У разі, коли всі респонденти позитивно оцінюють економічну ситуацію, $CCI = 200$; коли частки позитивних і негативних оцінок однакові, $CCI = 100$; коли переважають негативні оцінки, $CCI < 100$.

Для моніторингу ринкових тенденцій чи купівельної поведінки цільових груп споживачів стосовно окремих товарів або торговельних марок створюють спеціалізовані панелі. Наприклад, компанія Touchpoll Ukraine здійснює панельне дослідження споживчої поведінки представників преміум-класу (Elite Consumer Monitor) за такими категоріями: банківські послуги, автомобільний ринок, ринок преміального алкоголю, електроніка, ресторани.

У маркетингових дослідженнях аудиторії масової комунікації застосовують модифікації Consumer Scan — медіа-панелі, які ведуть спостереження за комунікативною поведінкою населення «з трьох екранів»: телебачення, інтернету і мобільних пристроїв. Методики медіаметричного вимірювання розвиваються за рахунок комерційних досліджень, провідні позиції в цій сфері посідає світова інформаційно-вимірвальна компанія Nielsen Media Research.

Об'єктами медіа-панелей слугують, як правило, домогосподарства. При формуванні панелі використовують низку параметрів, які визнані такими, що впливають на комунікативну поведінку аудиторії того чи іншого виду мас-медіа. Так, при формуванні ТВ-панелі такими параметрами є:

- технічні характеристики обладнання в домогосподарстві (тип приймання телесигналу, телевізійне обладнання, кількість телевізорів у домогосподарстві та ін.);

- соціально-демографічні характеристики домогосподарств (кількість осіб у родині; наявність дітей віком до 17 років; наявність людей віком понад 60 років; дохід домогосподарства, основна мова спілкування та ін.);

- соціально-демографічні характеристики кожного глядача (стать, вік, освіта, зайнятість на роботі).

Домогосподарства, що беруть участь в дослідженнях, добираються за спеціальною методикою. В Україні п'ятирічна ТВ-панель базується на загальнонаціональній вибірці; від 2014 року, обсяг її становить 3740 домогосподарств (без Криму), з них дві третини розташовані в містах із населенням понад 50 тис. осіб, третина — в містах із населенням менше 50 тис. осіб та в селах.

Завдяки тому, що обрані домогосподарства за своїми соціально-демографічними та технічними характеристиками відтворюють структуру населення України, вони формують розгорнену картину теле-аудиторії. Генеральну сукупність теле-аудиторії становлять глядачі віком від 4-х років.

Вимірювання теле-аудиторії здійснюють за допомогою спеціальних пристроїв електронного сканування комунікативної поведінки респондентів. У кожному домогосподарстві, що бере участь в дослідженні, до телевізора приєднують спеціальні пристрої — *people-metrics*, що фіксують час, упродовж якого відбувався перегляд телепрограм членами домогосподарства (віком від 4-х років). В Україні телеметричні дослідження організує ACNielsen. Результати дослідження використовують телеканали, рекламні агенції та рекламодавці.

У даний час посилюється інтерес до медіаметричних вимірювань інтернет-аудиторії. В Україні медіадослідження інтернет-аудиторії виконує компанія InMind Factum Group Ukraine. Дослідження ґрунтується на спостереженні постійної репрезентативної панелі інтернет-користувачів, тестування веб-активності яких здійснюється за допомогою спеціальних програм. Панель репрезентує інтернет-користувачів усієї країни віком 15 років і старше. За даними панельного моніторингу відвідуваності веб-сайтів, з одного боку, вивчається соціально-демографічний профіль інтернет-користувачів, а з іншого — оцінюються рейтинги інтернет-ресурсів.

Панельні обстеження набули широкого застосування в кон'юнктурних обстеженнях підприємств, головна мета яких — оцінювання стану ділової активності суб'єктів господарювання, аналіз причин можливих змін цього стану, відстежування циклів ділової активності та короткострокове прогнозування макроекономічних показників.

Домінантною формою статистичного спостереження в кон'юнктурних обстеження є постійні панелі. В Україні добір підприємств для обстеження, тобто формування панелі підприємств-респондентів, здійснюють за видами економічної діяльності (промисловість, будівництво, роздрібна торгівля, сільське господарство і транспорт) на підставі статистичного та адміністративного (ЄДРПОУ) реєстрів. Обсяг панелі не перевищує 10 % загальної кількості підприємств. За даними обстежень визначають: інтегральні (композиційні) індикатори для окремих видів

економічної діяльності, на їх основі — загальний (інтегральний) індикатор ділового клімату [Пугачева, 2007]. Гармонізація методології кон'юнктурних обстежень із системами Єврокомісії та ОЕСД дає змогу здійснювати міжнародні порівняння.

У вирішенні складних технічних, гуманітарних та соціально-економічних проблем широко застосовують експертні опитування. Експерти — це респонденти з необхідним рівнем компетентності щодо предмета чи об'єкта дослідження. Експертні опитування проводять спеціальні центри експертизи, наукові інформаційно-аналітичні центри й експертні групи, які створюються при державних установах і службах, а також на громадських засадах, у комерційних структурах, навчальні центри.

Експертне опитування вважається одним з найбільш раціональних способів вибору варіантів управлінських рішень, моделей поведінки в альтернативних ситуаціях, у визначенні пріоритетів, у діагностиці й прогнозуванні практично будь-яких ситуацій, у тому числі за умов неповної інформації. Водночас при формуванні експертних груп виникає проблема добору експертів, які б володіли необхідними для розв'язання досліджуваних проблем знаннями, досвідом, творчим потенціалом, умінням генерувати нові ідеї. Експертів багато не буває. Розмір вибірки зазвичай становить 15–20 експертів, тому репрезентації як такої у цьому виді спостереження не існує. Робота з експертами залежно від теми і конкретних завдань дослідження може бути організована у формі простого інтерв'ю, фокусування в групі, мозгового штурму, застосування методу Дельфі.

Крім класичних форм панельних обстежень, у практиці статистичних досліджень використовують інші форми панелей, зокрема омнібуси (omnibus panels). Ідея омнібуса полягає в тому, що спеціалізована компанія періодично (щомісяця, щокварталу) за репрезентативною вибіркою (у країні або регіоні) проводить багатоцільове обстеження, програма якого містить запропоновані різними замовниками запитання або блоки запитань з широкого спектра дослідницьких завдань. Цей вид панелі дає можливість кожному з учасників значно знизити витрати на проведення дослідження і в стислі терміни отримати результати разового польового етапу дослідження. Аби оцінити зміни, що відбуваються в досліджуваному процесі, необхідно повторно брати участь в омнібусі.

Вправи та запитання для самоконтролю

1. Поясніть суть і завдання повторних вибірових обстежень. За якими схемами можна організувати повторні обстеження?
2. Поясніть процедуру ротації вибірових одиниць.
3. Як визначити загальний обсяг ротаційної вибірки, який забезпечує надійну вибірову оцінку цільового показника?
4. Задля обстеження середньорічної заробітної плати у підприємствах із кількістю 10–49 осіб на регіональному рівні визначте, скільки звітних одиниць необхідно обстежити за умови, що середньорічна заробітна плата в генеральній сукупності таких підприємств становить 12,7 тис. грн при дисперсії 25,8 тис. Характеристика надійності результатів не має перевищувати 2,5 %. Величина дизайн-ефекту умовно приймається на рівні 1,5.
5. Як зменшити вплив нетипових одиниць на вибірові оцінки?
6. Поясніть особливості розміщення вибірки з питань праці за стратами (регіонами та видами економічної діяльності).
7. Розмістіть вибірку обсягом 400 домогосподарств за стратами:

Страта	Частка постійного населення, p_{wj}	Стандартне відхилення, s_{wj}
Великі міста	0,267	2106,5
Малі міста	0,360	1487,3
Райони	0,373	1898,5
Разом	1,000	X

8. Поясніть сутність композиційної оцінки поточного значення середньої з урахуванням ротації.
9. Як визначити оптимальну частку заміщення?
10. Поясніть взаємозв'язок між часткою збіжних одиниць у ротаційній вибірці і коефіцієнтом кореляції.
11. За даними ротаційної вибірки 40 малих підприємств громадського харчування у поточному періоді окупність витрат становила: по збіжній частині — 28 %, по незбіжній — 30 %. Частка заміщення становила 25 %, дисперсія окупності витрат по незбіжній частині сукупності в 1,25 рази більша. Визначте середній рівень окупності витрат на підприємствах громадського харчування в поточному періоді з урахуванням ротації. Якою була б оцінка середнього рівня окупності витрат, якби дисперсії в обох частинах сукупності були однаковими?

12. Проведено вибіркове обстеження туристичних агенцій у два раунди. Обсяг вибіркової сукупності $n = 10$, частка заміщення $(1 - d) = 30\%$. За результатами обстеження витрати на рекламу становили:

Період обстеження	Середні витрати, тис. гр. од.		Загалом по сукупності
	По збіжній частині	По незбіжній частині	
Попередній	3,1	2,8	3,0
Поточний	3,2	3,3	?

У поточному обстеженні дисперсія витрат на рекламу становила 2,9, коефіцієнт кореляції між значеннями витрат в агенціях, які обстежувалися в обох раундах, — 0,85, коефіцієнт регресії — 0,82. Методом регресії оцініть середні витрати на рекламу в поточному періоді з урахуванням ротації; визначте похибку вибірки для середніх витрат на рекламу.

13. За даними завдання 11 оцініть:

- виграш у точності ротаційної вибірки порівняно з повним заміщенням;
- зміну середніх витрат на рекламу в поточному періоді порівняно із попереднім;
- оптимальну частку заміщення сукупності при організації ротаційної вибірки, якщо коефіцієнти кореляції $r = 0,75$.

14. Проведено вибіркове обстеження учасників панельного опитування щодо споживчих настроїв населення. Обсяг вибіркової сукупності становив 150 респондентів, частка заміщення — 35%. За результатами обстеження індекс споживчих настроїв за бальною шкалою від 1 до 10 у двох послідовних раундах становив:

Період обстеження	Індекс споживчих настроїв		Загалом по сукупності
	По збіжній частині	По незбіжній частині	
Попередній	5,8	6,3	6,0
Поточний	6,0	6,2	?

Дисперсія оцінок споживчих настроїв у поточному обстеженні становила 1,4, коефіцієнт кореляції між оцінками респондентів, що брали участь в обох раундах, — 0,89, коефіцієнт регресії — 0,91.

Методом регресії оцініть значення індексу споживчих настроїв у поточному періоді з урахуванням ротації, визначте похибку вибірки для середнього значення індексу, а також оцініть зміну середніх.

16. За даними ротаційної вибірки 30 торговельних точок мережі супермаркетів (частка заміщення 23 %) приріст продажу безалкогольних напоїв у поточному періоді становив: по збіжній частині — 1,5 %, по незбіжній — 2,3 %. Дисперсія приросту продажу по незбіжній частині сукупності більша в 1,15 раза.

Визначте середній приріст продажу в мережі супермаркетів у поточному періоді з урахуванням ротації та похибку вибірки.

17. За результатами послідовних вибіркових обстежень 10 філій банку щодо суми наданих автокредитів необхідно визначити оцінку зміни середньої суми цих кредитів.

Філія банку	Травень	Червень	Філія банку	Травень	Червень
1	5,5		6	5,7	6,0
2	5,6		7	6,1	6,1
3	6,0	6,1	8	6,0	6,1
4	5,9	5,9	9		6,0
5	5,6	6,3	10		6,2

18. Поясніть загальний принцип організації панельних обстежень

19. Що є предметом дослідження торговельної панелі? Яким є принцип її формування і результати обстеження?

20. Поясніть принцип формування національної споживчої панелі.

21. Поясніть особливості медіа-панелей. Що є об'єктом дослідження у медіа-панелі?

22. Поясніть особливості вимірювання теле- та інтернет-аудиторії.

Рекомендована література

1. *Гладун О.М.* Вибіркові обстеження населення: методологія, методика, практика : Монографія. Ніжин: ТОВ «Видавництво «Аспект — Поліграф», 2008. —344 с.

2. Методика формування вибірки для проведення поточних обстежень підприємств із питань статистики праці / Державна служба статистики України. 2016. URL: www.ukrstat.gov.ua/metod_polog/metod_doc/.../metod.htm

3. Методологічні положення щодо формування вибіркових сукупностей для проведення органами державної статистики України вибіркових обстежень населення (домогосподарств). 2016. URL: zakon.rada.gov.ua/rada/file/text/62/f474085n19.pdf.

Розділ 10

ПОШИРЕННЯ РЕЗУЛЬТАТІВ ВИБІРКОВОГО ОБСТЕЖЕННЯ НА ГЕНЕРАЛЬНУ СУКУПНІСТЬ

Методичні засади поширення результатів вибірки на генеральну сукупність

Формування системи статистичних ваг для вибірок складного дизайну

Забезпечення надійності результатів обстеження на основі непрямих методів оцінювання

Після вивчення цього розділу студент має

- *знати*: способи відновлення пропущених даних і поширення результатів вибірки на генеральну сукупність; систему коефіцієнтів коригування ваг для вибірок складного дизайну; умови застосування методів непрямого оцінювання із використанням зовнішньої (додаткової) інформації;
- *уміти*: оцінювати стан самозваженості вибірки; задля забезпечення надійності результатів вибірки в умовах сучасного дизайну вибірки і технології опрацювання даних застосовувати найефективніші методи імпутації даних, системи вагових коефіцієнтів, комбінаційні оцінні функції.

Методичні засади поширення результатів вибірки на генеральну сукупність

Завершальний етап будь-якого вибіркового обстеження — поширення (екстраполяція) його результатів на генеральну сукупність. Головним методологічним принципом, на якому ґрунтується можливість використання результатів вибірки, є принцип репрезентативності, який вимагає незміщеності вибірових оцінок, наближення їх до дійсних значень параметрів генеральної сукупності.

Репрезентативність вибірки забезпечується двома формалізованими процедурами: 1) дизайном вибірки і 2) обсягом вибірки, який за обраного дизайну здатен забезпечити прийнятну точність

вибіркових оцінок. Нагадаємо, що за даними вибірки оцінюють, як правило, три параметри генеральної сукупності:

- середнє значення цільового показника (наприклад, середньодушові витрати домогосподарств на непродовольчі товари);
- частку одиниць сукупності з певним значенням ознаки (наприклад, частку домогосподарств, що мають 2-х і більше дітей);

- сумарне значення цільового показника (наприклад, доходи домогосподарств від підприємницької діяльності).

У разі, коли йдеться про середнє значення ознаки або частку одиниць сукупності з певним значенням ознаки, поширення вибіркових оцінок на генеральну сукупність зводиться до визначення того довірчого інтервалу, до якого з прийнятою ймовірністю може потрапити відповідний параметр генеральної сукупності (див. підрозділ 6.2). Щодо сумарного значення показника генеральної сукупності традиційно застосовують два способи поширення даних вибірки: пряме поширення (прямий перерахунок) і спосіб коефіцієнтів.

Прямий перерахунок можна здійснити на підставі вибіркової середньої або з використання частки вибірки (співвідношення чисельності вибіркової і генеральної сукупностей). У першому випадку вибіркова середня ознака \bar{x}_i помножується на кількість одиниць генеральної сукупності: $\sum_1 x_i = \bar{x}_i N$. Оскільки середня — величина випадкова, якій властива гранична похибка вибірки Δ_x , то і для сумарного показника в генеральній сукупності визначають довірчий інтервал:

$$(\bar{x} - \Delta_x^-)N \leq \sum_{i=1}^N x_i \leq (\bar{x} + \Delta_x^-)N.$$

Наприклад, загальна посівна площа під картоплею в господарствах населення в районі становить 2000 га. За даними вибіркового обстеження середня врожайність картоплі — 224,0 ц/га, гранична похибка середньої — 7,2 ц/га. Отже, можливий обсяг валового збору картоплі з цієї площі буде не меншим за 433,6 тис. ц [(2000 · (224,0 — 7,2))]. Максимальний валовий збір — 462,4 тис. ц [(2000 · (224,0 + 7,2))].

За другим способом сумарний показник генеральної сукупності $\sum_{i=1}^N x_i$ визначається множенням вибіркового сумарного значення $\sum_{i=1}^n x_i$ на коефіцієнт поширення $g = \frac{N}{n}$, який показує, скільки одиниць генеральної сукупності представляє одна вибіркова одиниця, або діленням на частку вибірки $d = \frac{n}{N}$, тобто:

$$\sum_{i=1}^N x_i = \sum_{i=1}^n x_i \frac{N}{n} = \sum_{i=1}^n x_i \frac{N}{n} .$$

Скажімо, за даними 2 % вибіркового обстеження підприємств малого бізнесу дебіторська заборгованість становила 216 млн грн. Звідси $g = \frac{1}{0,02} = 50$, а сума дебіторської заборгованості в генера-

льній сукупності $\frac{216}{0,02} = 216 \cdot 50 = 10\,800$ млн грн.

Спосіб коефіцієнтів застосовують також для уточнення даних суцільного спостереження щодо чисельності одиниць сукупності:

$$N^* = N \frac{n_1}{n_0} ,$$

де N^* та N — чисельності сукупності з поправкою і без поправки на недооблік;

n_0 та n_1 — чисельності в контрольних точках за первинними і уточненими даними.

Наприклад, після щорічного перепису худоби в господарствах населення проводиться 10 %-й вибірковий контроль, мета якого — визначити частку недообліку худоби. За даними перепису в районі налічується 10 200 корів. У домогосподарствах, які потрапили до контрольної вибірки, за переписом 200 корів, а за даними перевірки — 206. Отже, коефіцієнт, на який слід скоригувати результати перепису, становить $\frac{206}{200} = 1,03$, а скоригована чисельність корів — $10\,200 \cdot 1,03 = 10\,506$ корів.

Розглянуті способи поширення результатів вибірки на генеральну сукупність ґрунтуються на припущенні, що обстеження

виконане відповідно до процедури вибіркового методу, обстежено всі одиниці, які потрапили до вибірки, а вибіркові оцінки повною мірою відображають риси та властивості сукупності, з якої були обрані. Проте це припущення часто не виконується. Практично в усіх вибіркових обстеженнях до розряду актуальних потрапляє проблема недоотримання інформації (неповнота, пропуск даних). Традиційно поява пропусків зумовлюється низкою причин, серед яких: брак інформації з об'єктивних причин, некомпетентні відповіді на поставлені запитання, зумисне приховування інформації, втрати частини даних у процесі їх передавання чи зберігання тощо. Недоотримана з будь-яких причин інформація призводить до певного зміщення вибіркових оцінок і може суттєво впливати на надійність результатів вибіркового обстеження. Отже, постає проблема відновлення пропущених даних, і — що важливо — пошуку тих методик, за якими вони будуть відновлюватися.

У соціально-економічних дослідженнях масштаби неповних даних за різними типами обстежень коливаються в межах від 10 до 30 % залежно від природи самого об'єкта обстеження, способу обстеження, типу респондента, кваліфікації й сумлінності інтерв'юєра, привабливості програми опитувальників, часу проведення обстеження тощо. Забезпечити 100 % обстеження вибіркової сукупності доволі важко, але теорія вибіркового методу пропонує широкий спектр прийомів компенсації неповних даних.

Вирізняють два типи некомплектності даних: 1) часткові пропуски даних щодо окремих значень ознак і 2) повна відсутність даних щодо окремих одиниць вибіркової сукупності (неповне охоплення). У практиці вибіркових обстежень для компенсації недоотриманих даних застосовують:

- різні за складністю методики заміни відсутніх даних з використанням додаткової інформації (*imputation methods*);
- заміну відсутніх даних такими, які можна визначити методами математико-статистичного моделювання;
- коригування коефіцієнтів поширення даних вибірки на генеральну сукупність за допомогою системи ваг.

Процедура компенсації некомплектних вибіркових даних обирається на підставі попереднього контролю повноти масиву даних та з'ясування масштабів і механізму породження пропусків. Зауважимо, що видалення некомплектних об'єктів і аналіз

масиву даних «повного складу» можна застосувати лише за умови невеликої кількості пропусків, що об'єктивно звужує сферу використання такого підходу в соціально-економічних дослідженнях.

Розглянемо найпоширеніші методики імпутації даних. Сутність цієї процедури полягає в заміщенні відсутнього значення певною величиною (замінником). Таким замінником можуть бути:

1. Середнє значення ознаки x , обчислене за наявними даними вибіркової сукупності, в стратифікованій вибірці — середні значення окремих груп \bar{x}_j ; для дискретної або категоріальної ознаки — мода. Недоліком такого підходу є можливе викривлення розподілу даних і зменшення варіації, що впливає на якість вибіркових оцінок.

2. Значення ознаки, яке у складі вибірки найближче до цільового за визначеними критеріями (метод «hot deck»). Таке донорське значення можна підібрати як з усієї сукупності комплектних спостережень, так і з деякої підгрупи — кластера, до якого належить цільовий об'єкт.

3. Значення ознаки (донора) із зовнішнього джерела інформації (метод «cold-deck»). Заміщення пропусків даних із зовнішніх джерел, у тому числі з попередніх чи аналогічних обстежень, відбувається тоді, коли внутрішньої інформації виявляється недостатньо.

4. Значення, обчислене за допомогою регресії, побудованої за даними комплектних об'єктів (без пропусків). Якість такої заміни вочевидь залежить від вдалого вибору форми регресійної моделі.

Застосування більшості методик імпутації даних виправдане за дотримання умови missing at random, тобто коли пропуски випадкові. В іншому разі можливе викривлення розподілу даних і зменшення варіації, що вплине на якість вибіркових оцінок. Відношення кількості ампутуваних значень певної ознаки x до загальної кількості значень цієї ознаки (рівень імпутації) розглядають як одну із характеристик надійності результатів вибіркового обстеження.

Імпутацію методом «hot-deck» можна задіяти також для компенсації повних невідповідей, утворивши випадкову вибірку з комплектних одиниць. Добір донорів для заміщення відсутніх об'єктів здійснюють за єдиним списком основи вибірки з інтервалом $\frac{1}{\pi}$, де π — частка об'єктів, які не дали інформації, у загальному їх числі.

Наприклад, при обстеженні підприємств малого бізнесу із загальної сукупності 80 одиниць для обстеження дібрано 8, з них 2 одиниці не надали інформації, тобто частка некомплектних даних становить $\pi = 2 / 80 = 0,025$. Тоді добір одиниць для заміщення відсутніх слід здійснити з інтервалом $(1 : 0,025) = 40$. Початковий номер визначається за схемою випадкового добору.

Отже, шлях до успішного заміщення пропущених даних лежить у напрямі максимально можливого використання всієї наявної інформації щодо вибіркової сукупності, зокрема допоміжної інформації, отриманої з інших джерел, що підвищує стабільність результатів обстеження.

У стратифікованій вибірці заміну здійснюють лише в межах певних груп і підгруп основи вибірки, тобто замість пропущеної одиниці слід взяти рівноцінну їй за визначальними ознаками, скажімо, за територіальним розміщенням, за формою власності тощо.

Важливою умовою забезпечення репрезентативності стратифікованої вибірки є дотримання принципу пропорційного представництва окремих груп. У разі, коли через недоліки вибірки цей принцип порушено (різниця між частками окремих груп перевищує 3 %), аби уникнути систематичної похибки застосовують процедуру перезважування. Так, при визначенні середньої генеральної сукупності \bar{X} застосовують перезважування вибіркового групових середніх \bar{x}_j за фактичними частотами (частками) генеральної сукупності f_j :

$$\bar{X} = \sum_{j=1}^m \frac{N_j}{N} \times \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} = \sum f_j \bar{x}_j.$$

Інші способи відтворення пропорцій генеральної сукупності враховують частки недоотриманих даних в окремих групах. Щоб відновити початкові пропорції генеральної сукупності, здійснюють коригування вибірки або шляхом відсікання частини одиниць, частка яких у вибірці непропорційно велика порівняно з часткою в генеральній сукупності, або шляхом дублювання результатів спостереження за одиницями тих груп, які недостатньо широко представлені у вибірці.

Наприклад, при вивченні громадської думки стосовно якості наданих комунальних послуг сформована стратифікована вибірка обсягом $n = 400$ за схемою пропорційного добору. Першу групу становлять мешканці середмістя, другу — мешканці околиці. Через низку причин частина анкет була недоотримана, а частина забракована. Зрештою придатними для подальшого опрацювання виявилися 300 анкет, розподіл їх за групами респондентів наведено у табл. 10.1. Як свідчать дані, структура генеральної сукупності порушена. Частки придатних анкет за групами респондентів виявилися різними. Для відтворення пропорцій генеральної сукупності візьмемо однакову для обох груп частку придатних анкет — 75 %.

Таблиця 10.1

ВІДТВОРЕННЯ ПРОПОРЦІЙ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

Групи респондентів	Розподіл сукупності, % до підсумку		Отримано придатних анкет		Необхідно анкет для відтворення структури генеральної сукупності
	генеральної	випадкової	% до запланованих	кількість	
A	40	38	71	114	120
B	60	62	78	186	180
Загалом	100	100	X	300	300

Кількість анкет, необхідних для збереження представництва вибірки в масиві анкет для подальшого опрацювання, наведено в останній графі таблиці. Згідно з даними, в групі *A* з меншою часткою придатних анкет необхідно продублювати шість анкет, а в групі *B*, навпаки, шість анкет з подальшого опрацювання вилучити. Добір анкет для дублювання і вилучення здійснюється за процедурою випадкової вибірки.

Крім перелічених простих методів імпутації даних застосовують більш складні, зокрема: *Zet*-алгоритми, метод Бартлета, *resampling*-метод тощо. Мабуть, не існує універсального методу, що підходить для всіх ситуацій. Кожен варіант імпутації має плюси і мінуси. Корисно поєднувати кілька методів заміщення неповних даних у наперед визначений спосіб.

Формування системи статистичних ваг для вибірки складного дизайну

Системи статистичних ваг є одним із найбільш поширених і науково обґрунтованих інструментів забезпечення репрезентативності й поширення результатів вибірки на генеральну сукупність. Статистичною вагою називають кількість одиниць генеральної сукупності, які репрезентує одна одиниця вибірки. Вирізняють базову та кінцеву (повну) вагу. Визначальною є базова вага, тобто кількість одиниць генеральної сукупності, яку за дизайном має репрезентувати дібрана у вибірку одиниця. Базову вагу можна розглядати як величину, обернену до ймовірності добору. У математичному виразі, якщо одиниця сукупності може потрапити до вибірки з імовірністю p , то її базова вага буде $w = 1/p$.

Імовірність потрапити до власне випадкової чи систематичної вибірки для будь-якої одиниці генеральної сукупності визначається відношенням обсягу вибірки до обсягу генеральної сукупності, тобто часткою вибірки $f = \frac{n}{N}$. Одиниця, дібрана з імовірністю $1/20 = 0,20$, репрезентує 20 одиниць генеральної сукупності. У разі багатоступеневої вибірки ймовірність включення одиниці до вибірки акумулює ймовірності добору на всіх ступенях. Наприклад, за двоступеневої вибірки одиниця першого ступеня включається у вибірку з імовірністю p_1 , другого ступеня — з імовірністю p_2 , а базова ймовірність потрапити до вибірки для j -ої одиниці сукупності є добутком цих двох імовірностей:

$$p_j = p_1 \cdot p_2.$$

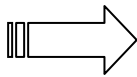
Припустімо, в одній сільській провінції з 80 сіл для обстеження дібрано 8, в кожному з них для обстеження добиралося кожне п'яте домогосподарство незалежно від їх кількості. Отже, загальна ймовірність потрапити до вибірки j -му домогосподарству становить: $p_j = 0,1 \times 0,2 = 0,02$. Відповідно, базова вага при поширенні результатів обстеження на генеральну сукупність буде $w_j = 1 / 0,02 = 50$.

Для ідеальної самозваженої вибірки, коли реалізуються всі процедури добору на всіх етапах обстеження, потреби в застосуванні системи ваг немає, оскільки ваги всіх одиниць вибірки однакові. На практиці від стану самозваженості найчастіше відхиляються вибірки зі складним дизайном (стратифіковані, бага-

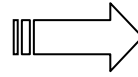
тоступеневі), зокрема вибірки населення (домогосподарств) через відмови домогосподарств від участі в обстеженні, наявність у списках для добору адрес неіснуючих домогосподарств, відхилення навантаження статистиків з проведення вибірових обстежень населення від теоретично визначеної величини тощо. Ці відхилення справляють різний вплив на систему ваг у різних територіальних одиницях. Для забезпечення репрезентативності результатів вибірок зі складним дизайном невіддільною складовою поширення результатів на генеральну сукупність є застосування системи ваг.

Саме шляхом зважування можна зменшити вплив на результати обстеження недосконалостей основи вибірки та окремих невибірових похибок. Важливою процедурою на етапі формування системи ваг є коригування базової ваги. Можна застосовувати кілька коригувань. Кількість одиниць генеральної сукупності, яку репрезентує обстежена вибіркова одиниця після коригувань, називають кінцевою вагою:

*Розрахунок
базової ваги*



*Коригування
базової ваги*



*Кінцева
вага*

Таким чином, система статистичних ваг відображає реальний дизайн вибірки і враховує основні етапи реалізації обстеження. При оцінюванні показників використовують, як правило, кінцеві ваги обстежених одиниць. Отже, коли в процесі проведення обстеження виникають певні дисбаланси вибірки, саме через систему ваг реалізуються процедури коригування (поліпшення) оцінок показників.

Методику формування системи статистичних ваг розглянемо на прикладі поширення результатів державного вибірового обстеження умов життя домогосподарств, яка розроблена відповідно до міжнародної практики проведення вибірових обстежень населення (домогосподарств).

Сучасний дизайн вибірового обстеження умов життя домогосподарств реалізує процедуру триступеневого добору: первинних територіальних одиниць, локальних (вторинних) територіальних одиниць і домогосподарств у межах локальних територіальних одиниць (див. підрозділ 8.1). При формуванні територіальної вибірки застосовують спосіб добору з імовірностями, пропорційними

її розміру (probability proportional to size, PPS). Характеристикою розміру територіальної одиниці є кількість домогосподарств. Головна мета застосування цієї процедури — забезпечити рівні ймовірності добору для кожної кінцевої одиниці вибірки. Ця

ймовірність має дорівнювати частці вибірки $f = \frac{n}{N}$.

Отже, ймовірність потрапити до вибірки для j -го домогосподарства залежить від ймовірності добору на всіх трьох ступенях формування вибірки, а базова вага j -го домогосподарства є оберненою величиною добутку цих ймовірностей:

По міських поселеннях По сільській місцевості

Загальна ймовірність
включення j -го домогосподарства до вибірки

$$p_j = p_1 \cdot p_2 \cdot p_3$$

$$p_j = p_1 \cdot p_2$$

Базова вага j -го домогосподарства

$$w_j = \frac{1}{p_j} = \left\{ \begin{array}{l} p_1 \cdot p_2 \cdot p_3 \\ 1 \quad 2 \quad 3 \end{array} \right.$$

$$w_j = \frac{1}{p_j} = \left\{ \begin{array}{l} p_1 \cdot p_2 \\ 1 \quad 2 \end{array} \right.$$

Розглянемо зміст і порядок розрахунку зазначених ймовірностей.

А. Міські поселення

1. Ймовірності добору первинних територіальних одиниць:

p_{1j} — ймовірність добору первинної територіальної одиниці (мале місто), у якому розташоване j -те домогосподарство ($j = 1, 2, \dots, n$). Самореферентивні міста добираються з ймовірністю $p = 1$, а для міст з чисельністю населення нижче порогу самореферентивності ймовірність пропорційна розміру територіальної одиниці:

$$p_{1j} = a \frac{M_a}{\sum M_a},$$

де a — кількість міських поселень, які треба дібрати в регіоні;

M_a — кількість домогосподарств у міському поселенні, для якого розраховується ймовірність p_{1j} ;

$\sum M_a$ — загальна кількість домогосподарств у міських поселеннях регіону.

2. *Ймовірності добору локальних (вторинних) територіальних одиниць:*

p_{2j} — ймовірність добору локальних територіальних одиниць вибірки розраховується за формулою:

$$p_{2j} = b \frac{M_b}{\sum M_b},$$

де b — кількість локальних територіальних одиниць, дібраних у конкретному міському поселенні;

M_b — кількість домогосподарств у локальній територіальній одиниці, для якої розраховується ймовірність p_{2j} ;

$\sum M_b$ — загальна кількість домогосподарства у дібраному міському поселенні ($\sum M_b = M_a$).

Якщо розміри локальних територіальних одиниць суттєво різняться в межах міста, може виникнути ситуація, коли ймовірність добору p_{2j} буде дорівнювати або перевищувати одиницю. У такому разі ймовірність приймається на рівні «одиниці» і локальна територіальна одиниця враховується як саморепрезентативна, відповідно коригуються кількість одиниць b і кількість домогосподарств M_b .

Загальна ймовірність включення до вибірки локальної територіальної одиниці визначається як добуток імовірностей ($p_{1j} \cdot p_{2j}$).

3. *Ймовірності добору домогосподарств*

p_{3j} — ймовірність добору домогосподарств у межах локальної територіальної одиниці характеризує шанс конкретного j -го домогосподарства бути дібраним:

$$p_{3j} = \frac{n_j}{M_b} = \frac{f}{(p_{1j} \cdot p_{2j})},$$

де n_j — кількість домогосподарств, які необхідно дібрати в межах локальної територіальної одиниці;

M_b — кількість домогосподарств у локальній територіальній одиниці.

За таких умов базова вага однакова для всіх домогосподарств і визначається за формулою:

$$w_j = \frac{1}{p_{1j} \cdot p_{2j} \cdot p_{3j}} = \frac{1}{f}.$$

В. Сільська місцевість

1. Ймовірності добору первинних територіальних одиниць:

p_{1j} — ймовірність включення до вибірки одиниць першого ступеня добору в сільській місцевості (сільських рад) по кожному району розраховується за формулою:

$$p_{1j} = C \frac{M_c}{\sum M_c},$$

де c — кількість дібраних сільських рад у районі;

M_c — кількість домогосподарств у дібраній сільській раді, для якої розраховується p_{1j} ;

$\sum M_c$ — загальна кількість домогосподарств у сільській раді.

2. Ймовірності добору домогосподарств:

p_{2j} — теоретична ймовірність добору j -го домогосподарства дібраної сільської ради розраховується за формулою:

$$p_{2j} = \frac{n_j}{M_c} = \frac{f}{p_{1j}}.$$

За таких умов базова вага однакова для всіх домогосподарств:

$$w_j = \frac{1}{p_{1j} \cdot p_{2j}} = \frac{1}{f}.$$

Базові ваги розраховуються на весь термін експлуатації вибірки. Для територіальної вибірки домогосподарств термін експлуатації становить п'ять років, тобто впродовж цього періоду дизайн і обсяг вибірки не змінюються.

Кінцевий масив первинних даних, отриманих за результатами вибіркового обстеження, містить, як правило, інформацію не за дібраними одиницями, а за одиницями, які вдалося обстежити. Фактичні ймовірності добору домогосподарств в межах територіальних одиниць можуть відрізнятися від базових через наявність зачинених і нежитлових приміщень, тривалу відсутність членів домогосподарств вдома, відмови окремих домогосподарств від участі в обстеженні тощо. Відхилення, що виникають в процесі реалізації обстеження, знижують надійність вибірових оцінок і потребують відповідного коригування.

Ураховувати ефекти впливу різних факторів на кінцеві ваги домогосподарств має система коефіцієнтів коригування ваг. Кінцева вага j -го домогосподарства w_j розраховується як добуток базової ваги домогосподарства w_{bj} та відповідних коефіцієнтів коригування ваг: (коефіцієнтів переважування):

$$w_j = w_{bj} \cdot k_{1j} \cdot k_{2j} \dots \cdot k_{mj},$$

де k_{1j} k_{2j} ... k_{mj} — коефіцієнти коригування ваг; m — загальна кількість етапів коригування, зокрема:

k_{1j} — коефіцієнт врахування нежитлових та зачинених приміщень;

k_{2j} — коефіцієнт врахування відмов домогосподарств від обстеження;

k_{2j} коефіцієнт врахування відмов окремих осіб від обстеження;

k_{2j} коефіцієнт компенсації постстратифікаційних ефектів.

Таким чином, через систему статистичних ваг реалізуються різні процедури коригування (поліпшення) оцінок показників, тим самим зменшується вплив на результати обстеження недосконалоостей основи вибірки. Статистичне зважування:

– підвищує рівень репрезентативності вибірки в плані відображення нею кількісних та структурних характеристик генеральної сукупності;

– дозволяє ефективно вирішувати такі проблеми як: урахування нерівних ймовірностей добору одиниць вибірки, фактичних рівнів відмов суб'єктів від участі в обстеженні, узгодження результатів обстеження з достовірною зовнішньою інформацією.

З метою наближення результатів вибіркового обстеження до наявної зовнішньої інформації та підвищення рівня надійності оцінок показників, які вимірюються шляхом вибіркового спостереження, до статистичних ваг застосовують процедуру калібрації, яка передбачає коригування системи ваг з метою узгодження певних показників обстеження з інформацією зовнішніх джерел. В обстеженні рівня життя населення (домогосподарств) зовнішні джерела представлені даними:

– демографічної статистики щодо чисельності постійного населення на початок звітного року за регіонами, типами місцевості та статеві-вікової структури населення;

– статистичної звітності щодо населення, яке перебуває в інституціональних закладах.

Забезпечення надійності результатів обстеження на основі методів непрямого оцінювання

Надійність вибірових оцінок є одним з основних вимірів якості результатів вибірового обстеження за обраного дизайну і обсягу вибірки. За своїм змістом надійність є характеристикою, зворотною точності, тобто чим менша похибка вибірки, тим вища надійність статистичної оцінки і навпаки. Необхідну надійність/точність вибірових даних забезпечують великі за обсягом сукупності із поділом такої сукупності на складові. За значної варіації ознак надійність/точність прямих вибірових оцінок знижуватиметься. Так, у річному структурному обстеженні малих підприємств надійність показників є достатньою на національному рівні, на рівні регіонів та видів економічної діяльності (згідно з КВЕД), але за видами економічної діяльності в регіонах виявляється значно нижчою, що пояснюється відносно малими обсягами вибірки й обмеженими можливостями оптимізації дизайну вибірки.

Проблема недостатньої надійності прямого оцінювання показників на рівні малих територій, адміністративних об'єктів (окремі міста, сільські райони) або так званих малих доменів (таких як види економічної діяльності, верстви населення, етнічні групи тощо) часто виявляється недостатньою і для уточнення потребує застосування спеціальних методів, серед яких за умов сучасного дизайну вибірки й технології опрацювання даних найефективнішими визнано методи непрямого оцінювання.

Характерною рисою непрямих методів оцінювання є, крім безпосередніх результатів обстеження, залучення з допоміжних джерел інформації щодо ознак, які в обстеженні не піддаються прямому вимірюванню. Це можуть бути дані попередніх етапів обстеження, кореляти цільових показників, отримані з адміністративних джерел, даних статистичної звітності, переписів, інших обстежень. За певних умов при однаковому обсязі вибірки за допомогою непрямих методів можна забезпечити вищу надійність вибірових даних порівняно з прямим оцінюванням. Необхідною (але недостатньою) умовою такої ситуації є наявність зв'язку між даними вибірки і даними з допоміжних джерел інформації.

Ці дані поєднуються за шляхом так званого «аналізу спряженості», згідно з яким дані вибірки у розглядаються як певна фун-

кція допоміжних (чи попередніх) даних x , тобто $Y = f(x)$. Очевидно, що та частина варіації ознаки y , яка зумовлена зв'язком з x , не є випадковою, а отже, при розрахунку похибок вибірки цю частину варіації необхідно виключити. До непрямих методів, які можуть підвищити точність вибіркової оцінки цільової ознаки y завдяки використанню з допоміжних джерел інформації щодо взаємозв'язаної з нею ознаки x , належать, зокрема, методи оцінювання за відношенням і лінійне оцінювання за регресією.

Оцінювання методом відношень (*ratio*) застосовують до відносних величин, які характеризують структуру сукупності, динаміку явища, інтенсивність поширення явища в середовищі. Скажімо, при оцінюванні частки банківських кредитів під заставу житла в загальному обсязі кредитів, частки витрат домогосподарств на оплату житлово-комунальних послуг, частки жінок серед трудових мігрантів тощо. Параметр генеральної сукупності, що підлягає оцінюванню, можна подати відношенням:

$$k = \frac{\sum y}{\sum x} = \frac{y}{\bar{x}}, \text{ або } y = kx.$$

Вибірковий розподіл відношення k більш складний, ніж розподіл середніх чисельника y і знаменника x , які змінюється від вибірки до вибірки. Для невеликих за обсягом вибірок розподіл відношень асиметричний і його оцінка має певне зміщення.

Математичне сподівання відношення k має вигляд:

$$E(k) = \frac{1}{k} \left[1 + \frac{(N-n)}{Nn} (V_x^2 - rV_x V_y) \right],$$

де V_x і V_y — коефіцієнти варіації ознак x і y , відповідно;

r — коефіцієнт кореляції між ознаками x і y .

Як міру точності оцінювання за відношенням (*ratio*) за аналогією з простими функціями використовують похибки вибірки:

- стандартну $\mu_{rat} = \sqrt{\frac{1-f}{n} s_w^2}$;
- відносну $\mu_{rat} = \frac{\mu_{rat}}{k}$.

Розрахунок похибки вибірки для відношення середніх k ґрунтується на тій частині варіації, яка розглядається як випадкова,

тобто на відхиленнях $(y - kx)$. Дисперсія цих відхилень s_w^2 визначається за формулою:

$$s_{(y-kx)}^2 = s_w^2 = s_y^2 + s_x^2 k^2 - 2rs_y s_x k.$$

Для зручності розрахунків стандартної похибки дисперсії ознак x і y замінюють квадратами коефіцієнтів варіації відповідно v_y^2 і v_x^2 . Тоді без поправки на скінченність сукупності:

$$\mu_{rat}^2 = k^2 \frac{1}{n} (V_y^2 + V_x^2 - 2rV_y V_x);$$

$$\mu_{rat}^2 = \frac{1}{n} (V_y^2 + V_x^2 - 2rV_y V_x).$$

Оцінка за відношенням зміщена. Величина зміщення (*bias*) залежить передусім від обсягу вибірки та щільності зв'язку. Її можна визначити за формулою:

$$bias(k) = \frac{y}{x} \left(\frac{N-n}{Nn} \right) (V_x^2 - rV_y V_x),$$

де V_x і V_y — коефіцієнти варіації ознак x і y відповідно;

r — коефіцієнт кореляції як міра щільності зв'язку між x і y .

Як видно з формули, зміщення оцінки відношення середніх k зменшується зі збільшенням обсягу і частки вибірки. Зі збільшенням обсягу вибірки розподіл оцінок наближається до нормального, а зміщення оцінки зменшується. На величину зміщення впливає також щільність зв'язку між ознаками x та y : чим більше значення коефіцієнта кореляції r , тим меншим є зміщення оцінки.

Наприклад, для визначення відносної зміни витрат на комунальні послуги два роки проводилося обстеження одних і тих самих 100 домогосподарств ($f = 0,008$). Якщо середні витрати на комунальні послуги за перший рік позначити \bar{x} , за другий — \bar{y} , то параметром, який підлягає оцінюванню, буде індекс витрат $\frac{\bar{y}}{\bar{x}}$

Припустімо, індекс витрат $\frac{\bar{y}}{\bar{x}} = 1,2$, варіація витрат не змінилася:

$V_y = V_x = 1,8$, $r = 0,78$. Тоді зміщення дорівнює:

$$\text{bias} \left(\frac{y}{x} \right) = 1,2 (1/100) (1,8^2 - 0,78 \times 1,8 \times 1,8) = 0,00856,$$

що у процентах до вибіркової оцінки індексу витрат (1,2) становить:

$$100 \frac{0,00856}{1,2} = 0,713\%.$$

Отже, для вибірок такого обсягу оцінка за відношенням має тенденцію завищувати реальне значення показника приблизно на 0,7 %. Для більших обсягів сукупності величина зміщення ще менша і її часто ігнорують.

Виграш у точності оцінки за відношенням порівняно з оцінкою простої випадкової вибірки визначають так:

$$\gamma_{\text{rat}} = \frac{\mu^2(\bar{y})}{\mu^2(\bar{y}/\bar{x})} = \frac{v_y^2 + v_x^2 - 2rv}{v_x^2 + v_y^2 - 2rv} = \frac{1}{1 + (V_x/V_y)(V_x/V_y - 2r)}.$$

Як видно з формули, точність оцінки за відношенням залежить від варіації ознак x і y та щільності зв'язку між ними. Виграш у точності оцінки за відношенням можливий за умови, що $r > \frac{V_x}{2V_y}$. У разі, коли $V_x = V_y$ (однаковий рівень варіації ознак), ефективність оцінки за відношенням становить:

$$\gamma_{\text{rat}} = \frac{1}{2(1-r)}.$$

У разі, коли між ознаками x і y існує кореляційний зв'язок, застосовують непрямий метод оцінювання за регресією, сутність якого — зв'язати дані вибіркового обстеження з результатами інших обстежень, щоб виключити з варіації ту її частину, яка зумовлена зв'язком. Поділ варіації y на випадкову s_W^2 і невипадкову s_B^2 складові здійснюється за правилом декомпозиції варіації:

$$s_y^2 = s_B^2 + s_W^2 = s_y^2 r^2 + s_y^2 (1 - r^2),$$

де s_y^2 — загальна дисперсія y ;

r — коефіцієнт кореляції.

Перший член цього рівняння $s_B^2 = s_y^2 r^2$ характеризує варіацію показника y «по лінії регресії», другий $s_W^2 = s_y^2(1 - r^2)$ — варіацію «навколо регресії», тобто випадкову (залишкову) варіацію⁵. Оцінюючи точність вибіркових характеристик, виключають варіацію «по лінії регресії» й використовують лише випадкову дисперсію, що дає істотний вигравш у точності оцінок.

Для будь-якої конкретної вибірки обсягом n похибка оцінки за регресією визначається наближено:

$$\mu_{regr}^2 \cong \left(\frac{N-n}{Nn} \right) s_y^2 (1 - r^2) \left[1 + \frac{1}{n} \right].$$

Без поправки на скінченність сукупності:

$$\mu_{regr}^2 = \frac{1}{n} s_y^2 (1 - r^2).$$

Логіку оцінювання методом регресії розглянемо на прикладі безперервного статистичного контролю якості металопродукції, ґрунтованим на кореляційному зв'язку між механічними властивостями і хімічним складом листового прокату. Обсяг вибірки — 2500 плавок.

Залежність тимчасового опору листового прокату y від вуглецевого еквіваленту x описується лінійним рівнянням регресії:

$$Y = 712,35 + 17,5x.$$

Коефіцієнт кореляції становить $r = 0,75$, що свідчить про достатньо високу щільність зв'язку між цими показниками.

За результатами вибірки середнє значення вуглецевого еквіваленту $\bar{x} = 2,14$ %, дисперсія тимчасового опору листового прокату за допоміжними даними — 128. І якщо виключити з варіації тимчасового опору прокату ту частину, яка залежить від хімічно-

⁵ У методі регресії формула залишкової дисперсії має такий вид: $s_w^2 = s_y^2 + s_x^2 b^2 - 2rs_y s_x b$. Якщо в правій частині рівності замінити b на вираз $r \frac{s_y}{s_x}$,

то $s_w^2 = s_y^2(1 - r^2)$.

го складу, це дасть можливість уточнити похибку вибірки й довірчий інтервал механічних властивостей листового прокату.

Використовуючи параметри лінійного рівняння, визначимо середній рівень тимчасового опору прокату (в мегапаскалях) за регресією:

$$\bar{y} = 712,35 + 17,5 \times 2,14 = 749,8,$$

а також похибку вибірки для середнього рівня. Оскільки обсяг вибірки великий ($n = 2500$), розрахунок здійснюється без поправки на скінченність сукупності:

$$\mu^2 = \frac{1}{n} s_y^2 (1 - r^2) = \frac{1}{2500} 128 (1 - 0,75^2) = 0,0224 .$$

Звідси $\mu = \sqrt{0,0224} = 0,15$, а довірчі межі середнього рівня $749,8 \pm 2 \times 0,15$.

Розмір похибки вибірки визначений на основі залишкового компоненту дисперсії $s_y^2 (1 - r^2)$. Якби такого уточнення не було, то квадрат похибки становив би $\mu^2 = \frac{1}{2500} 128 = 0,0512$, або в 2,3 раза більше.

Виграш у точності регресійної оцінки порівняно з простою випадковою можна визначити відношенням:

$$\gamma_{regr} = \frac{\mu^2}{\mu_{regr}^2} = \frac{1}{1 - r^2} = \frac{1}{1 - 0,75^2} = \frac{1}{0,4375} = 2,285 .$$

Отже, при $r \neq 0$, незалежно від напрямку зв'язку, регресійна оцінка завжди буде ефективнішою за оцінку простого випадкового добору, оскільки множник $(1 - r^2) < 1$. За відсутності зв'язку, коли $r = 0$, похибка μ буде похибкою простої випадкової вибірки, а в разі, коли $r \rightarrow 1$, похибка $\mu \rightarrow 0$.

Залежність оцінок за регресією і за відношенням від коефіцієнта кореляції r і ефективність їх щодо оцінки простої випадкової вибірки можна простежити за даними табл. 10.2 [Джессен, 1895: с. 133].

Таблиця 10.2

**ВІДНОСНА ЕФЕКТИВНІСТЬ ОЦІНОК ПО ВІДНОШЕННЮ І ПО РЕГРЕСІЇ
ПОРІВНЯНО З ОЦІНКОЮ ПРОСТОЇ ВИПАДКОВОЇ ВИБІРКИ**

Коефіцієнт кореляції, r	Відносна ефективність оцінок, γ	
	Оцінка за відношенням $1/[2(1-r)]$	Оцінка за регресією $1/(1-r^2)$
-1,0	0,25	∞
-0,9	0,263	5,27
-0,75	0,2857	2,28
-0,50	0,33	1,33
-0,25	0,40	1,067
0	0,50	1,0
0,25	0,67	1,067
0,50	1,0	1,33
0,75	2,0	2,28
0,90	5,0	5,27
1,0	∞	∞

Як свідчать дані таблиці, оцінка за відношенням точніша за оцінку простої випадкової вибірки лише при $r > 0,5$, тоді як оцінка за регресією «безпечна», тобто ризик втрати точності відсутній за будь-якої щільності зв'язку, незалежно від його напрямку. Отже, якщо допоміжна інформація містить ознаку, яка корелює з об'єктом дослідження y , оцінювання методом регресії може дати значний вигравш в точності. У разі лінійної регресії y на x зміщення оцінки дорівнює нулю.

Оцінювання методом регресії виявляється ефективним у багатозадачних вибірках, коли вибіркові оцінки однієї фази використовують як додаткову інформацію на наступних фазах, що підвищує точність результатів вибіркового обстеження й зменшує витрати на обстеження.

Розглянемо методику уточнення вибірових оцінок методом регресії на прикладі двофазного (*подвійного*) випадкового добору. Припустімо, що із первинної сукупності дібрано n' одиниць, за якими вимірюється лише ознака x . Із сукупності n' одиниць, дібраних на першій фазі, добирається n'' одиниць другої фази, в

рамках якої вимірюється ознака y . Оцінка середнього значення \bar{y} задається формулою:

$$\bar{y} = \bar{y}' + b'(\bar{x}' - \bar{x}'),$$

де \bar{x}' і \bar{y}' — середні значення ознак у вибірці другої фази;

\bar{x} — середнє значення ознаки x у вибірці першої фази;

b'' — коефіцієнт регресії y на x за даними вибірки другої фази.

Оцінка середньої двофазної (*double*) вибірки дуже схожа на регресійну оцінку, але замість середнього значення \bar{x} для первинної сукупності використовують його оцінку \bar{x}' за даними вибірки першої фази. Похибка середньої \bar{y} визначається за формулою:

$$\mu_{\bar{y}}^2 = \frac{N - n'}{N} \left[\frac{s^2(1 - r^2)}{n'} \left(1 + \frac{n' - 1}{n'} \times \frac{1}{n'} \right) \right] + \frac{r^2 s_y^2}{n'}.$$

Якщо не враховувати скінченність сукупності і відношення $\frac{n' - n'}{n'} \times \frac{1}{n' - 3}$, то формула набуває вигляду:

$$\mu_{\bar{y}}^2 = \frac{r^2 s_y^2}{n'} + \frac{(1 - r^2) s_y^2}{n'}.$$

Коли у двофазній вибірці вимірювання взаємопов'язаних ознак x і y здійснюється на різних фазах, виникає питання, яку частину із наявних ресурсів спрямувати на вимірювання x , а яку — на вимірювання y ? Для оптимального вибору n' і n'' розглянемо просту функцію витрат.

$$C = n'c' + n''c'',$$

де c' і c'' — одиничні витрати на вимірювання ознаки x у першій фазі і ознаки y — у другій фазі, а C — допустимі сумарні витрати.

Похибка вибірки буде наближатися до мінімального рівня тоді, коли:

$$\frac{n'}{n''} = \sqrt{\frac{1 - r^2}{r^2} \times \frac{c'}{c''}}.$$

У випадку, коли сумарні витрати C задані, а розподіл їх між двома фазами оптимальний, можна визначити відносну ефективність двофазного добору порівняно з одинарним (простим випад-

ковим). Оскільки для одинарного добору $C = nc''$, а $\mu_y^2 = \frac{s_y^2}{C/c'}$, то відносна ефективність двофазної вибірки з урахуванням витрат на обстеження буде:

$$\gamma_{doub} = \frac{c'}{(\sqrt{(1-r^2)c'} + r\sqrt{c'})^2}.$$

Загалом двофазний добір ефективніший за простий випадковий тоді, коли:

$$\frac{c'}{c'} > \frac{r^2}{(1-\sqrt{1-r^2})^2}, \text{ або } r^2 > \frac{4c'c'}{(c' + c')^2},$$

тобто у тих випадках, коли:

- витрати на вимірювання ознаки x набагато менші порівняно з витратами на вимірювання y ;
- кореляція між x і y достатньо висока.

У табл. 10.3 наведено співвідношення одиничних витрат, які забезпечують задані значення відносної ефективності двофазного добору порівняно з однофазним при різних коефіцієнтах кореляції [Джессен, 1985: с. 306].

Таблиця 10.3

**ОПТИМАЛЬНІ СПІВВІДНОШЕННЯ ОДИНИЧНИХ ВИТРАТ
ПРИ ДВОФАЗНОМУ ДОБОРІ**

Коефіцієнт кореляції, r	Відношення одиничних витрат c''/c' , необхідне для забезпечення відносної ефективності на рівні			
	1,00	1,25	1,50	2,00
0,65	7,33	23,36	132,04	X
0,70	6,00	15,07	46,77	∞ (0,707)
0,75	4,91	10,36	23,40	X
0,80	4,00	7,38	13,65	55,79
0,85	3,23	5,35	8,61	22,22
0,90	2,55	3,85	5,59	11,01
0,95	1,91	2,66	3,55	5,79
1,0	1,00	1,25	1,50	2,00

Наприклад, для оцінки середньої страхової суми на страхування життя у використаємо оцінку за регресією. Страхова сума корелює з доходами сімей x . Відповідні дані:

N	C	c''	c'	r	b''
5000	100	1,0	0,1	0,9	3,9

Оптимальне розміщення при двофазному доборі визначимо за рівністю:

$$n' = n' \sqrt{\frac{1-0,9^2}{0,9^2}} \times \frac{0,10}{1,00} = 0,153n'.$$

Звідси: $C = c' n' + c'' 0,153 n'$;

$$n' = \frac{100}{0,10 + 1,00 \times 0,153} = 395;$$

$$n'' = 0,153 \times 395 = 60.$$

Отже, найефективнішим буде розміщення вибірки між двома фазами:

- вибірка першої фази, в рамках якої визначаються доходи сімей, $n' = 395$;
- вибірка другої фази, де реєструються дані про страхові суми, $n'' = 60$.

Щоб відповісти на питання, якому добору (двофазному чи простому випадковому) віддати перевагу при вказаних одиничних витратах на обстеження, визначимо відносну точність і відносну ефективність двофазного добору γ_{doub} . Про ефективність двофазного добору свідчить співвідношення одиничних витрат і щільності зв'язку:

$$0,9^2 > \frac{4 \times 1,00 \times 0,10}{(1,00 + 0,10)^2}, 0,81 > 0,33.$$

Відносна точність двофазного добору щодо простого випадкового становить:

$$\gamma_{doub} = \frac{1,00}{\left[\sqrt{(1-0,81) \times 1,00} + 0,9 \sqrt{0,10} \right]^2} = \frac{1,00}{(0,435 + 0,284)^2} = 1,94,$$

тобто при заданих витратах точність двофазної вибірки майже вдвічі перевищує точність однофазної вибірки.

У практиці вибірових обстежень двофазна вибірка часом поєднується зі стратифікованою. У таких випадках за даними обстеження вибірки першої фази сукупність поділяється на групи, а формування вибірки другої фази здійснюється за принципами стратифікованої вибірки.

У сучасній статистичній практиці непрямі методи оцінювання стають ефективним інструментом розв'язання проблеми підвищення надійності оцінок показників на різних рівнях агрегації даних. Нині зростає актуальність оцінювання показників доходів, витрат, бідності, зайнятості й безробіття, умов життя і стану здоров'я окремих верств населення. В процесі моніторингу стану соціально-економічної ситуації й оцінювання ефективності соціальних програм акценти зміщуються на регіональний та локальний (місцевий) рівень, відтак, зростає потреба в статистично надійній інформації в регіональному та районному розрізах, на рівні муніципалітетів, соціальних груп населення тощо.

Серед традиційних методів «малих територій» і «малих домогів», які пропонуються для застосовування в офіційній статистиці для вирішення проблеми недостатньої надійності оцінок показників, пріоритетне значення має метод композиційного оцінювання (composite estimation or shrinkage estimation) [Статистика підприємства, 2017]. Історично метод композиційного оцінювання виник у результаті пошуку можливостей об'єднання (композиції) прямих і непрямих оцінних функцій. Пряме оцінювання показників здійснюється за стандартними прямими оцінними функціями, тобто оцінка показника розраховується на підставі поточних даних вибіркового обстеження. Як непрямі можуть застосовуватися так звані синтетичні оцінні функції, побудовані на основі статистичних моделей з використанням зовнішньої (додаткової) інформації.

Отже, композиційна оцінка для певного цільового показника Y^C — це функція, отримана шляхом об'єднання прямої і непрямой оцінок. У спрощеному вигляді така оцінка має вигляд:

$$Y^C = \varphi Y^{DIR} + (1 - \varphi) Y^{NDIR},$$

де Y^{DIR} — пряма оцінка (оцінка Гурвіца-Томпсона);

Y^{NDIR} — непряма оцінка;

φ — ваговий коефіцієнт, який визначається за умови мінімізації стандартної похибки композиційної оцінки ($\mu(Y^C)$):

$$\mu(Y^C) = \varphi^2 \mu(Y^{DIR}) + (1 - \varphi)^2 \mu(Y^{NDIR}) \rightarrow \min,$$

$$\varphi = \frac{MSE(Y^{NDIR})}{MSE(Y^{DIR}) + MSE(Y^{NDIR})},$$

де $MSE(Y^{DIR})$ — стандартна похибка прямої оцінки;

$\mu(Y^{NDIR})$ — стандартна похибка непрямої оцінки.

Отже, у практиці вибірових обстежень для забезпечення необхідної надійності оцінок цільових показників застосовують різні методологічні підходи, зокрема:

- обґрунтування дизайну вибірки з урахуванням вимог щодо необхідного рівня надійності оцінок;
- використання ускладнених методів опрацювання первинних даних (імпутації пропуску даних, спеціальних процедур опрацювання екстремальних значень тощо);
- застосування досконалих процедур зважування та методів калібрації статистичних ваг для врахування наявної зовнішньої інформації;
- застосування спеціальних методів оцінювання показників, зокрема на основі методів моделювання;
- застосування процедур об'єднання даних з різних джерел.

Вправи та запитання для самоконтролю

1. Як поширюються дані вибірки на генеральну сукупність при визначенні сумарних величин?

2. В одному з районів області зареєстровано 300 торговельних кіосків і 200 аптечних. Задля перевірки даних щодо розвитку мережі підприємств роздрібної торгівлі обрано контрольні пункти, в яких зареєстровано 50 торговельних і 40 аптечних кіосків. За результатами перевірки виявлено: торговельних кіосків — 54, аптечних — 42. Визначте коефіцієнт недообліку і кількість підприємств роздрібної торгівлі з поправкою на недооблік.

3. За даними обстеження 75 домогосподарств поселення витрати на послуги охорони здоров'я та придбання медичних препаратів становив-

ли в середньому на одне домогосподарство 120 грн на місяць. Частка обстежених домогосподарств 10 %. Визначте загальний обсяг витрат у генеральній сукупності домогосподарств.

4. Які способи компенсації неповноти даних застосовують при поширенні результатів вибірки на генеральну сукупність? Дайте характеристику способу середніх і відношень.

5. За яких умов для компенсації неповноти даних застосовують кластерні процедури?

6. Поясніть сутність переважування. Як саме формується система ваг?

7. При формуванні вибіркової сукупності обстеження умов життя домогосподарств обсяг вибірки з урахуванням рівня надійності результатів обстеження становив 12900, загальна кількість неінституціональних домогосподарств — 17610 тис. Визначте базовий коефіцієнт поширення результатів вибірки на генеральну сукупність (базову вагу). Скоригуйте коефіцієнт поширення, якщо рівень відмов домогосподарств від обстеження становить 4 %.

8. Поясніть концептуальні засади непрямого оцінювання. Чи доцільно застосовувати методи непрямого оцінювання за відсутності зв'язку між даними вибірки і зовнішніх джерел інформації?

9. Вкажіть, який метод непрямого оцінювання — за регресією чи за відношенням — дасть вигреш в точності, якщо зв'язок між ознаками y і x характеризується коефіцієнтом кореляції $r = +0,50$.

10. Поясніть, чому оцінка за регресією безпечніша порівняно з оцінкою за відношенням. Яким буде вигреш у точності при $r = +0,8$?

11. За даними вибіркового обстеження 30 домогосподарств пенсіонерів-одинаків маємо:

№ з/п	Сукупні витрати за місяць, грн	Витрати на комунально-побутові послуги, грн
1	1890	378
2	2026	422
...
Загалом	56350	11268
Коефіцієнт варіації	0,85	0,46

За умови, що коефіцієнт кореляції між сукупними витратами і витратами на комунально-побутові послуги становить $r = 0,6$, визначте:

- частку витрат на комунально-побутові послуги;
- стандартну похибку і довірчі межі для частки.

12. За даними вибіркового обстеження 20 емітентів, акції яких у лютому були виставлені на торги на фондовій біржі, середня вартість реалізованих акцій становила 28,5 млн грн при дисперсії 79. Зв'язок між вартістю реалізованих акцій y і статутним фондом емітентів x описується рівнянням регресії $Y = 10,4 + 0,24 x$, коефіцієнт кореляції $r = 0,7$. Середній розмір статутного фонду підприємств-емітентів — 160 млн грн.

Визначте похибку вибірки і довірчі межі для середньої вартості реалізованих акцій.

13. Кошторис вибіркового обстеження становить $C = 1200$ грн. Одиначні витрати на обстеження на першій фазі $c' = 0,1$; на другій — $c'' = 0,8$. Коефіцієнт кореляції між ознаками $r = 0,90$.

Доведіть, який добір ефективніший — одиничний чи подвійний? За умови організації подвійної (двофазної) вибірки визначте оптимальне розміщення вибірки за фазами.

14. На замовлення компанії-виробника газованих напоїв біля виходу з магазину впродовж тижня проводилося опитування споживачів з метою виявлення частки осіб, які споживають солодкі газовані напої. Опитування здійснювалося методом двофазного добору: на першій фазі людей запитували, чи споживають вони солодкі газовані напої взагалі (ознака $x = 0;1$), а на другій фазі — чи купували вони такий напій упродовж останнього тижня (ознака $y = 0;1$). Перше запитання ставили всім респондентам, тоді як друге лише опитаним на сьомий, тобто останній день проведення обстеження. Було отримано такі результати опитування:

1-а фаза	$n' = 843$	$\bar{x}' = 0,20$	
2-га фаза	$n' = 367$	$\bar{x}' = 0,33$	$\bar{y}' = 0,45$

Дисперсії ознак становлять $s_x^2 = 0,17$; $s_y^2 = 0,23$, коефіцієнт регресії $b' = 0,46$, коефіцієнт кореляції $r = 0,61$. Визначте методом регресії частку людей, які споживають солодкі газовані напої, а також похибку вибірки для частки.

Оцініть виграш у точності оцінки методом регресії порівняно з простою випадковою вибіркою, якщо одиначні витрати на обстеження становлять: на першій фазі $c' = 0,5$ гр. од., на другій — $c'' = 0,7$ гр. од.

Визначте оптимальне співвідношення кількості опитаних на першій і другій фазах, яке забезпечить мінімальну похибку вибірки.

Рекомендована література

1. *Гладун О.М., Сариогло В.Г.* Напрямки підвищення якості даних вибіркового обстеження умов життя домогосподарств та вирішення проблеми „малих територій” // *Статистика України.* — 2003. — № 1. — С. 4–11.
2. *Гончар О.В.* Забезпечення якості статистичної інформації: методологія та організація : Монографія. К. : ВПД «Формат», 2011. — 372 с.
3. *Кузнєцова Н.В.* Виявлення та оброблення невизначеностей у формі неповних даних методами інтелектуального аналізу // *Системні дослідження та інформаційні технології.* — 2016. — № 2. — С. 104–115. URL: dspace.nbuv.gov.ua/handle/123456789/134018
4. *Лиса О.В.* Оцінювання показників робочої сили за даними щомісячних обстежень населення // *Статистика України.* — 2008. — № 2 (41). — С. 76–81.
5. *Сариогло В.Г.* Проблеми статистичного зважування вибірових даних : Монографія. К. : ІВЦ Держкомстату України, 2005. — 264 с.
6. *Сариогло В.Г.* Оцінювання соціально-економічних показників: прикладні аспекти застосування непрямих методів : Монографія. К. : Інститут демографії та соціальних досліджень ім. М.В. Птухи НАН України, 2012. — 136 с.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Балабанов Г. В., Вишневецький В.В. Територіальний соціально-економічний моніторинг в Україні. — Київ: Ін-т географії НАН України, Нора-Друк, 2001. — 58 с.
2. Валенте П. Проведение переписей в Европе: как считают население в раунде 2010 года // Вопросы статистики. — 2012, №12. — С.3–8.
3. Васечко О.О. Методологічні основи статистики підприємств : Монографія. — К. : НТК стат. дослідж., 2005. — 296 с.
4. Вибіркове спостереження: Термінол. словн. / Укл.: О.О. Васечко, О.І. Черняк, Є.М. Жуйкова, О.В. Гончар та ін. — К. : Держкомстат України, 2004. — 140 с.
5. Волков А.Г. Из истории переписи населения 1937 года // Вестник статистики. — 1990, № 8. — С. 45–56.
6. Галіцин В. К. Системи моніторингу : Монографія. — К. : КНЕУ, 2000. — 232 с.
7. Гладун О.М. Вибіркові обстеження населення: методологія, методика, практика : Монографія. — Ніжин: ТОВ «Видавництво «Аспект – Поліграф», 2008. — 344 с.
8. Гладун О.В., Ващаєва Н.А. Дизайн вибірки та проблеми його оптимізації // Проблеми теорії та методології бухгалтерського обліку, контролю і аналізу : Зб. наук. пр. — Житомир: ЖДТУ, 2007. — Вип. 2(8). — С. 31–39.
9. Гладун О.М., Саріогло В.Г. Напрямки підвищення якості даних вибіркового обстеження умов життя домогосподарств та вирішення проблеми «малих територій» // Статистика України. — 2003. — № 1. — С. 4–11.
10. Гончар О.В. Забезпечення якості статистичної інформації: методологія та організація : Монографія. — К. : ВПД «Формат», 2011. — 372 с.
11. Джессен Р. Методы статистических обследований. — М. : Финансы и статистика, 1985. — 478 с.
12. Дружинин Н.К. Выборочный метод и его применение в социально-экономических исследованиях. — М. : Статистика, 1970. — 104 с.
13. Дружинин Н.К. Выборочное наблюдение и эксперимент. Москва: Статистика, 1977. — 175 с.
14. Єріна А.М. Організація вибірових обстежень : Навч. посіб. — К. : КНЕУ, 2003. — 127 с.
15. Єріна А.М., Пальян З.О. Статистика : Підруч. — К. : КНЕУ, 2010. — 351 с.
16. Закон України «Про інформацію». 1991. URL: <http://ukrstat.gov.ua>
17. Збарская И.А. Перепись населения: вызовы и перспективы раунда 2020 года // Вопросы статистики. — 2012. — № 10. — С. 9–13.

18. *Кокрен У.* Методы выборочного обследования. — М. : Статистика, 1976. — 440 с.
19. Концептуальні основи статистичного моніторингу : Монографія / Ра ред. М.В. Пугачової. — К. : ІВЦ Держкомстату України, 2003. — 343 с.
20. *Кузнецова Н.В.* Виявлення та оброблення невизначеностей у формі неповних даних методами інтелектуального аналізу // Системні дослідження та інформаційні технології. — 2016. — № 2. — С. 104–115. URL: dspace.nbuv.gov.ua/handle/123456789/134018
21. *Лиса О.В.* Оцінювання показників робочої сили за даними щомісячних обстежень населення // Статистика України. — 2008. — № 2 (41). — С. 76–81.
22. Методика формування вибірки для проведення поточних обстежень підприємств із питань статистики праці / Державна служба статистики України. 2006. URL: www.ukrstat.gov.ua/metod_polog/metod_doc/.../metod.htm
23. Методика формування вибіркової сукупності звітних одиниць у структурному обстеженні малих підприємств / Державна служба статистики України. 2010. URL: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2010/536/metod.htm
24. Методика формування вибірових сукупностей для проведення у 2014–2018 роках вибірових обстежень населення (домогосподарств): умов життя домогосподарств, економічної активності населення та сільськогосподарської діяльності населення у сільській місцевості / Державна служба статистики України. 2013. URL: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2013/37/metod_vyb_suk.zip
- Методологічні положення щодо підготовки звітів з якості даних річного вибірового структурного обстеження малих підприємств / Державна служба статистики України. 2010. URL: www.ukrstat.gov.ua/metod_polog/metod_doc/.../metod.htm
26. Методологічні положення щодо формування вибірових сукупностей для проведення органами державної статистики України вибірових обстежень населення (домогосподарств) / Державна служба статистики України. 2018. URL: zakon.rada.gov.ua/rada/file/text/62/f474085n19.pdf.
27. Методологічні положення з організації державного статистичного спостереження щодо структурних змін в економіці України та її регіонів / Державна служба статистики України. 2016. URL: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2016/134/134_2016.htm
28. *Пальян З. О.* Демографія : Навч. посіб. — К. : ВПЦ «Київський університет», 2014. — 222 с.
29. *Пархоменко В.М.* Методи вибірових обстежень : Навч. посіб. — К. : КНУ ім. Тараса Шевченка, 2001. — 148 с.

30. Перший Всеукраїнський перепис населення: історичні, методологічні, соціальні, економічні, етнічні аспекти / За ред. І.Ф. Кураса, С. І.Пирожкова. — К. : ІОЦ Держстату України, 2004. — 558 с.
31. Положення про Реєстр статистичних одиниць та формування основ вибірки сукупостей одиниць статистичних спостережень за діяльністю підприємств / Державна служба статистики України. 2009. URL: <http://www.ukrstat.gov.ua/edrpoj/prso.htm>
32. Пугачова М.В. Методологічні засади статистичного моніторингу ділової активності підприємств : Монографія. — К. : ДП Інформ-анал. агентство, 2007. — 535 с.
33. Рекомендации Конференции европейских статистиков по проведению переписей населения и жилищного фонда в 2020 году. — Нью-Йорк, Женева: ООН, ECE/ces/2015/41. — 259 с.
34. Саріогло В.Г. Проблеми статистичного зважування вибірових даних : Монографія. — К. : ІВЦ Держкомстату України, 2005. — 264 с.
35. Саріогло В.Г. Оцінювання соціально-економічних показників: прикладні аспекти застосування непрямих методів : Монографія. — К. : Інститут демографії та соціальних досліджень імені М.В. Птухи НАН України, 2012. — 136 с.
36. Системи фінансового моніторингу. Методологія проектування / За ред. О.В. Мозенкова. — Харків: ВД «ІНЖЕК», 2005. — 151 с.
37. Составление планов выборки для обследований домашних хозяйств: практические рекомендации. — ООН, Нью-Йорк: 2010. — 256 с.
38. Статистика підприємств: навч. посіб. / За ред. С.О. Матковського. — К. : Алерта, 2017. — 560 с.
39. Стеценко С. Г. Демографічна статистика : Підручник. — К. : Вища школа, 2005. — 415 с.
40. Черняк О.І. Техніка вибірових досліджень. — К. : МІВВІЦ, 2001. — 248 с.
41. Шварц Д. Выборочный метод. Руководство по применению статистических методов оценивания / Пер. с нем. — М. : Статистика, 1978. — 213 с.
42. Швець В., Царук О. Методологічні засади побудови системи статистичного моніторингу державного боргу // Вісник КНУ ім. Тараса Шевченка. Економіка. — 2008. — Вип. 99–100. — С. 56–59.
43. Юл Дж., Кендалл М. Теорія статистики. — М. : Госстатиздат, 1960.
44. Kish L. Survey sampling. — Wiley Classics Library Edition, 1995. — 643 p.
45. Palian Z.O., Tishchenko A.O. Modern political views of Ukrainian population: statistical evaluation and analysis // Models and methods for analysing and forecasting economic processes. Theory and practice. Cracow University of Economics Press, 2016. — P. 110–119. URL: http://plicki.konferencjajakopianska.pl/proceedings_2016/pdf/Palian_Tischenko.pdf

ДОДАТКИ

Додаток А

Приклад використання таблиці випадкових чисел

Таблиця випадкових чисел містить серії цифр, які дібрані за допомогою електронних сигналів і чергуються випадково. Для більшої наочності і зручності використання цифри розміщені по 5. Оскільки кожна цифра і їх послідовність випадкові, то таблицею можна користуватися, переміщуючись або по вертикалі або по горизонталі.

Припустимо, із сукупності 700 одиниць необхідно відібрати 35. Початкову точку в таблиці вибирають навмання. Нехай такою точкою буде перший стовпчик шостого рядка. Оскільки обсяг генеральної сукупності виражається трьох значним числом, то вибираються лише три останні цифри, значенням яких менше 700. Рухаючись по рядку, визначимо номери, які будуть включені у вибірку, а саме: 238, 126, 306, 114, наступний код пропускаємо, оскільки $718 > 700$, 584, 291, 575, 075 і т. д. Процедура продовжується доти, доки кількість дібраних номерів не сягне необхідного обсягу вибірки ($n=35$).

Можна було б використати і перші три цифри, тоді у вибірку потрапили б одиниці з іншими номерами.

Фрагмент таблиці випадкових чисел

44 983	33834	54280	67850	96025	96117	00768	14821
89 494	34431	44890	59892	79682	20308	82510	53609
54430	52632	94126	95597	48338	67645	44676	14730
96999	42104	34377	63309	82181	00278	28209	95629
87947	09427	32380	43636	58578	07761	28456	46570
30238	46126	85306	37114	22718	50584	92291	56575
22938	13073	32066	43098	75738	94910	15403	89151
89182	27750	63314	87302	49472	24885	79506	60638

Додаток Б

Додаток Б 1

Стандартні похибки параметрів генеральної сукупності за умови нормального розподілу (Юл, Кендал, с. 479, 512-513)

Параметр	Стандартна похибка	Параметр	Стандартна похибка
Частка	$\frac{pq}{\sqrt{n}}$	Центральний момент третього порядку	$\sigma^3 \sqrt{\frac{6}{n}}$
Середня	$\frac{\sigma}{\sqrt{n}}$	Центральний момент четвертого порядку	$\sigma^4 \sqrt{\frac{96}{n}}$
Дисперсія	$\sigma^2 \sqrt{\frac{2}{n}}$	Медіана	$1,25331 \frac{\sigma}{\sqrt{n}}$
Середнє квадратичне відхилення	$\frac{\sigma}{\sqrt{2n}}$	Децилі 4 і 6	$1,26804 \frac{\sigma}{\sqrt{n}}$
Коефіцієнт варіації	$\frac{V}{\sqrt{2n}} \sqrt{1 + \frac{2V^2}{10^4}}$	Децилі 3 і 7	$1,31800 \frac{\sigma}{\sqrt{n}}$
Коефіцієнт регресії	$\frac{\sigma_1 \sqrt{1-r^2}}{\sigma_2 \sqrt{n}}$	Децилі 2 і 8	$1,42877 \frac{\sigma}{\sqrt{n}}$
Коефіцієнт кореляції Пірсона	$\frac{1-r^2}{\sqrt{n}}$	Децилі 1 і 9	$1,70942 \frac{\sigma}{\sqrt{n}}$
Коефіцієнт кореляції рангів Спірмена	$\frac{1-\rho^2}{\sqrt{n}}$	Квартилі	$1,36263 \frac{\sigma}{\sqrt{n}}$

Додаток Б 2

**Формули оцінювання параметрів
при простому випадковому доборі**

Параметр сукупності	Вибіркова оцінка
Сумарне значення ознаки	$X = \frac{N}{n} \sum_{i=1}^n x_i$
Середнє значення	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Дисперсія ознаки	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
Стандартна похибка оцінки сумарного значення ознаки	$\mu_Y = \frac{Ns}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$
Стандартна похибка оцінки середнього значення ознаки	$\mu_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$
Відносна похибка вибірки	$V_{\mu} = \frac{\mu}{\bar{x}} 100$

В таблиці використані позначення:

i — номер одиниці сукупності ($i = 1, 2, 3, \dots, N$ або $i = 1, 2, 3, \dots, n$)

N — обсяг генеральної сукупності,

n — обсяг вибіркової сукупності,

x_i — значення ознаки у i -ої одиниці сукупності.

Додаток Б 3

Формули оцінювання параметрів при стратифікованій вибірці

Параметр сукупності	Вибіркова оцінка
Сумарне значення ознаки	$X = \sum_1^m N_j \bar{x}_j$
Середнє значення	$\bar{x} = \frac{1}{N} \sum_1^m N_j \bar{x}_j$
Дисперсія ознаки для j -ої страти (групи)	$s_j^2 = \frac{1}{n_j - 1} \sum_1^{n_j} (x_{ij} - \bar{x}_j)^2$
Середня з групових дисперсій (випадкова)	$s_w^2 = \frac{\sum_1^m \sum_1^{n_j} (x_{ij} - \bar{x}_j)^2}{n - m}$
Стандартна похибка оцінки сумарного значення ознаки	$\mu_Y = \sqrt{\sum_1^m N_j (N_j - n_j) \frac{s_j^2}{n_j}}$
Стандартна похибка оцінки середнього значення ознаки	$\mu_{\bar{x}} = \sqrt{\frac{1}{N^2} \sum_1^m N_j (N_j - n_j) \frac{s_j^2}{n_j}}$
Відносна похибка вибірки	$V = \frac{\mu}{\bar{x}} 100$

В таблиці використані позначення:

i — номер одиниці сукупності ($i = 1, 2, 3, \dots, N$ або $i = 1, 2, 3, \dots, n$);

j — номер групи ($j = 1, 2, 3, \dots, m$);

m — кількість груп;

N_j — обсяг j -ої групи в генеральній сукупності;

n_j — обсяг j -ої групи у вибірковій сукупності;

x_{ij} — значення ознаки у i -ої одиниці сукупності, яка належить до j -ої групи;

\bar{x}_j — середнє значення ознаки у j -ій групі.

Додаток Б 4

Формули оцінювання параметрів при кластерній вибірці

Параметр сукупності	Вибіркова оцінка
Сумарне значення ознаки	$X = \frac{M}{m} \sum_1^m X_i$
Середнє значення ознаки по гніздах	$\bar{X} = \frac{1}{m} \sum_1^m X_i$
Середнє значення ознаки	$\bar{x} = \frac{X}{N}$
Стандартна похибка оцінки сумарного значення ознаки	$\mu_Y = \sqrt{M^2 \left(1 - \frac{m}{M}\right) \frac{1}{m} \frac{1}{m-1} \sum_1^m (X_i - \bar{X})^2}$
Стандартна похибка оцінки середнього значення ознаки	$\mu_{\bar{x}} = \frac{1}{N} \sqrt{M^2 \left(1 - \frac{m}{M}\right) \frac{1}{m} \frac{1}{m-1} \sum_1^m (X_i - \bar{X})^2}$
Відносна похибка вибірки	$V = \frac{\mu}{\bar{x}} 100$

В таблиці використані позначення:

j — номер гнізда ($j = 1, 2, 3, \dots, m$);

m — кількість гнізд у вибірці;

M — кількість гнізд у генеральній сукупності;

X_j — сумарне значення ознаки у j -ому гнізді;

\bar{X} — середнє значення ознаки по гніздах.

Додаток Б 5

Формули оцінювання параметрів при двоступеневій вибірці

Параметр сукупності	Вибіркова оцінка
Сумарне значення ознаки	$X = \frac{M}{m} \sum_{i=1}^m \frac{N_i}{n_i} \sum_{j=1}^{n_j} x_{ij} ; X_i = \frac{N_i}{n_i} \sum_{j=1}^{n_j} x_{ij} ;$
Кількість одиниць в генеральній сукупності	$N = \frac{M}{m} \sum_{i=1}^m N_i$
Середнє значення ознаки	$\bar{x} = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_j} x_{ij} ; \bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_j} x_{ij}$
Кількість одиниць в середньому на групу	$n = \frac{1}{m} \sum_{i=1}^m n_i$
Стандартна похибка оцінки сумарного значення ознаки	$\mu_X = \sqrt{\frac{K_1 s_B^2 + \sum_{i=1}^m K_2 s_{Wi}^2}{1}}$
Стандартна похибка оцінки середнього значення ознаки	$\mu_{\bar{x}} = \frac{1}{N} \sqrt{K_1 s_B^2 + \sum_{i=1}^m K_2 s_{Wi}^2}$
Відносна похибка вибірки	$V_{\mu} = \frac{\mu}{\bar{x}} 100$

В таблиці використані позначення:

i — номер одиниці сукупності ($i = 1, 2, 3, \dots, N$ або $i = 1, 2, 3, \dots, n$);

j — номер групи ($j = 1, 2, 3, \dots, m$);

M — кількість груп в генеральній сукупності;

m — кількість дібраних груп;

N_j — обсяг j -ої групи в генеральній сукупності;

n_j — кількість дібраних елементів у j -ій групі;

x_{ij} — значення ознаки у i -ої одиниці сукупності, яка належить до j -ої групи;

\bar{x}_j — середнє значення ознаки у j -ій групі;

s_B^2 — між групова дисперсія;

s_{Wi}^2 — групова дисперсія;

$s_{\bar{w}}^2$ — дисперсія середня з групових.

Додатки В

Додаток В 1

Квантилі t -розподілу Стьюдента $t_{1-\alpha}(k)$

k	$\alpha = 0,025$	$\alpha = 0,05$	k	$\alpha = 0,025$	$\alpha = 0,05$
4	2,78	2,13	16	2,12	1,75
5	2,57	2,01	18	2,10	1,73
6	2,45	1,94	20	2,09	1,73
7	2,37	1,89	22	2,07	1,72
8	2,31	1,86	24	2,06	1,71
9	2,26	1,83	26	2,06	1,71
10	2,23	1,81	28	2,05	1,70
11	2,20	1,80	30	2,04	1,70
12	2,18	1,78	40	2,02	1,68
13	2,16	1,77	50	2,01	1,68
14	2,15	1,76	100	1,98	1,66
15	2,13	1,75	∞	1,96	1,64

Додаток В 2

Коефіцієнти співвідношення варіаційного розмаху і середньоквадратичного відхилення для різного обсягу сукупності ($\sigma \approx kR$)

n	10	20	30	40	50	100	200
k	0,32	0,27	0,24	0,23	0,22	0,20	0,18

Додаток В 3

**Індикатори якості статистичних продуктів, рекомендовані
Євростатом для відображення у звітах з якості**




№ з/п	Компонент якості	Індикатор якості / Назва
1	Відповідність / Релевантність	Рівень доступності результатів спостереження, Індекс задоволеності користувача, Проміжок часу, що минув від останнього опитування користувача
2	Точність / Надійність	Коефіцієнт варіації Рівень над охоплення Рівень неузгодженості при редагуванні Рівень відповідей Рівень повних відповідей Рівень імпутації Кількість помилок за типами Середній обсяг перевірок
3	Своєчасність і пунктуальність	Проміжок часу від закінчення звітного періоду та датою оприлюднення перших результатів Проміжок часу від закінчення звітного періоду та датою оприлюднення остаточних результатів Пунктуальність публікацій
4	Доступність і зрозумілість / Ясність	Кількість розсилок та/або проданих публікацій за видами Кількість відвідувань онлайн баз даних Рівень повноти метаданих
5	Послідовність / Порівняльність Зіставність / Узгодженість	Довжини порівняльних часових рядів Асиметрія показників дзеркальної статистики

Продовження додатку Г

2. Житлові умови домогосподарства

+

+ (питання 1-7 на тимчасово проживаючих, проживаючих в інституціональних установах та безпритульних не заповнюються)

1. Тип житлового приміщення	1 <input type="checkbox"/> Індивідуальний будинок 2 <input type="checkbox"/> Частина індивідуального будинку 3 <input type="checkbox"/> Окрема квартира 4 <input type="checkbox"/> Загальна (комунальна) квартира 5 <input type="checkbox"/> Гуртожиток 6 <input type="checkbox"/> Готель 7 <input type="checkbox"/> Інше приміщення 8 <input type="checkbox"/> Наймане житлове приміщення у окремих громадян
2. Кількість зайнятих житлових кімнат	 житлових кімнат 1 <input type="checkbox"/> частина кімнати
3. Розмір площі індивідуального будинку, частини індивідуального будинку, окремої або загальної (комунальної) квартири (кв. м)	 загальна +  житлова
4. Ви або члени Вашого домогосподарства маєте земельні ділянки сільськогосподарського призначення	1 <input type="checkbox"/> так 2 <input type="checkbox"/> ні
5. Якщо так, то на яких правах	1 <input type="checkbox"/> на правах приватної власності 2 <input type="checkbox"/> на правах оренди 3 <input type="checkbox"/> інше
6. Яка загальна площа Ваших земельних ділянок	1 <input type="checkbox"/> менше 0,1 га 4 <input type="checkbox"/> від 1,1 до 5 га 2 <input type="checkbox"/> від 0,1 до 0,5 га 5 <input type="checkbox"/> від 5,1 до 10 га 3 <input type="checkbox"/> від 0,51 до 1 га 6 <input type="checkbox"/> більше 10 га
7. Чи здасте Ви свої земельні ділянки в оренду	1 <input type="checkbox"/> так 2 <input type="checkbox"/> ні

+

Продовження додатку Г

Зміст відповідей заповненої перенесеної документації зберігається пошкоджені будь-кому. Ці дані використовуються лише для статистичних цілей у зведеному знеособненому вигляді.

ДЕРЖАВНИЙ КОМІТЕТ СТАТИСТИКИ УКРАЇНИ
ВСЕУКРАЇНСЬКИЙ ПЕРЕПИС
НАСЕЛЕННЯ 2001 р.
РЕПИСНИЙ ЛИСТ

Форма 2С
ЗАТВЕРДЖЕНО
Наказ Держкомстату
України
30.03.2001 № 175
Зразки нанесення міток

		1		2			
№ перелисного відділу	№ інструкторської дільниці	№ лінійної дільниці	№ приміщення у межах лінійної дільниці	список проживаючих у приміщенні членів домогосподарства №	№ а.п. особи у межах домогосподарства	№ бланка у межах портфеля	№ портфеля
2С							
Прізвище та ініціали							
1. Ваші родинні стосунки з особою, що записана першою у домогосподарстві				10. Ваша освіта (для осіб віком 6 років і старших)			
1 <input type="checkbox"/> член домогосподарства, записаний першим	4 <input type="checkbox"/> мати, батько	7 <input type="checkbox"/> свекрука, свекор, тесть, теща	10 <input type="checkbox"/> інший ступінь споріднення та спорідняті	1 <input type="checkbox"/> повна вища (лицей)	6 <input type="checkbox"/> базова загальна середня (початкова)	7 <input type="checkbox"/> початкова загальна (початкова)	8 <input type="checkbox"/> не між початкової загальної
2 <input type="checkbox"/> дружина, чоловік	5 <input type="checkbox"/> сестра, брат	8 <input type="checkbox"/> баба, дід	11 <input type="checkbox"/> не родич	3 <input type="checkbox"/> початкова вища (середня-спеціальна)	4 <input type="checkbox"/> незакінчена вища (для осіб, які закінчили навчання в 1996 році / раніше)	9 <input type="checkbox"/> повна загальна середня (середня загальна)	9 <input type="checkbox"/> неписемний(на)
3 <input type="checkbox"/> дочка, син	6 <input type="checkbox"/> невістка, зять	9 <input type="checkbox"/> онука, онук	№ матері (батька)		Навчальний заклад, який закінчив(ла), скільки класів (курсів), в якому році		
2. Ваша стать				11. Ви закінчили професійно-технічний навчальний заклад (для осіб віком 15 років і старших)			
1 <input type="checkbox"/> чоловіча 2 <input type="checkbox"/> жіноча				1 <input type="checkbox"/> так 2 <input type="checkbox"/> ні			
3. Тимчасово				12. Тип навчального закладу, в якому Ви навчаєтесь (для осіб віком 6 років і старших)			
1 <input type="checkbox"/> відсутній 2 <input type="checkbox"/> проживаючий				1 <input type="checkbox"/> вищий 3 <input type="checkbox"/> загальноосвітній 5 <input type="checkbox"/> не навчаєтесь			
4. Ваша дата народження				2 <input type="checkbox"/> професійно-технічний 4 <input type="checkbox"/> інший навчальний заклад (курс)			
число	місяць	рік	скільки виповнилося років	Для осіб віком до 7 років, які не відвідують школу, казати, чи відвідують дошкільний заклад 6 <input type="checkbox"/> так 7 <input type="checkbox"/> ні			
				13. Ваші джерела засобів існування			
5. Ваше місце народження (вказіть: Автономна Республіка Крим, область, м.м. Київ, Севастополь, інша держава — назва)				1 <input type="checkbox"/> робота на підприємстві, в організації, установі, селянському (фермерському) господарстві			
				2 <input type="checkbox"/> робота у окремих громадах			
6. Ваше етнічне походження (вказіть: національність, (народність) або етнічну групу)				3 <input type="checkbox"/> робота на вільному підприємстві			
				4 <input type="checkbox"/> робота на індивідуальній основі			
7. Ваші мовні ознаки				5 <input type="checkbox"/> робота у вільному селянському (фермерському) господарстві			
а) рідна мова (вказіть)				6 <input type="checkbox"/> робота на сімейному підприємстві без оплати праці			
б) якщо Ваша рідна мова не українська, то вкажіть чи володієте вільно українською мовою 1 <input type="checkbox"/> так 2 <input type="checkbox"/> ні				7 <input type="checkbox"/> робота в особистому підсобному господарстві			
в) інша мова, якою Ви вільно володієте (вказіть)				8 <input type="checkbox"/> прибуток від власності			
				9 <input type="checkbox"/> пенсія			
8. Ваше громадянство				10 <input type="checkbox"/> стипендія			
1 <input type="checkbox"/> Україна 2 <input type="checkbox"/> без громадянства				11 <input type="checkbox"/> допомога (крім допомоги по безробіттю)			
інша держава — назва				12 <input type="checkbox"/> допомога по безробіттю			
				13 <input type="checkbox"/> інший вид державного забезпечення			
9. Ваш сімейний стан (для осіб віком 15 років і старших)				14 <input type="checkbox"/> на утриманні інших осіб			
1 <input type="checkbox"/> ніколи не перебував(ла) у шлюбі	5 <input type="checkbox"/> розлучений(ла)	6 <input type="checkbox"/> розділений(лася)	номер основного джерела засобів існування				
2 <input type="checkbox"/> перебуваю у зареєстрованому шлюбі	№ дружини (чоловіка)						
3 <input type="checkbox"/> перебуваю у незареєстрованому шлюбі							
4 <input type="checkbox"/> удівець, удова							
				15 <input type="checkbox"/> інше джерело			

Продовження додатку Г

Питання 14-19 на тимчасово проживаючих не заповнюються

14. Повна назва установи, організації, підприємства, власної справи, де Ви зайняті основною роботою, вид діяльності (вказіть)

15. Ваше заняття на основній роботі

+

(вказіть докладну назву професії, посади або роботи, що виконується Вами)

16. Місцезнаходження Вашої роботи

на території проживання

+ на іншій території (вказіть)

(Автономна Республіка Крим, область, м.м. Київ, Севастополь, інша держава – назва)

17. Ваше положення у занятті (ким Ви працюєте)

1 працівнич(а) за наймом 3 роботодавець 5 безоплатно працюючий член сім'ї

2 член колективного підприємства, кооперативу 4 самозайнят(а)

18. У цьому населеному пункті Ви проживаєте безперервно з народження

1 так 2 ні

а) якщо «ні», то вкажіть рік, з якого безперервно проживаєте

час переїзду припадає на період з 5.12.2000 до 5.12.2001 3

б) місце попереднього проживання (вказіть)

(Автономна Республіка Крим, область, м.м. Київ, Севастополь, інша держава – назва)

в) тип населеного пункту, з якого Ви прибули

4 міський 5 сільський +

г) чи Ви є:

6 біженець 7 депортований

19. Ви народжували дітей *Для жінок (віком 15 років і старших)*

так ні

якщо «так», вкажіть

+ а) скільки дітей народили б) скільки з них живі +

Продовження додатку Д

2

ОБСТЕЖЕННЯ СІЛЬСЬКОГОСПОДАРСЬКОЇ ДІЯЛЬНОСТІ НАСЕЛЕННЯ В СІЛЬСЬКІЙ МІСЦЕВОСТІ

2. Скільки всього осіб в цей час мешкає у Вашому домогосподарстві?

10

УВАГА, ФАХІВЦЮ З ІНТЕРВ'ЮВАННЯ!

Коди для заповнення питання 1 даного розділу:

гр. 3 "Родинні стосунки з членом домогосподарства, записаним під №1"	
Член домогосподарства, записаний першим – голова домогосподарства	1
Дружина, чоловік	2
Дочка, син	3
Мати, батько	4
Сестра, брат	5
Невістка, зять	6
Свекруха, свекор, теща, тесть	7
Баба, дід	8
Онук, онука	9
Інша ступінь споріднення та своєїства	10
Не родич	11
Невідповідь	99

гр.8 "Коди рівня освіти (для осіб 6 років і старше)"	
Повна вища	1
Базова вища	2
Неповна вища	3
Повна загальна середня	4
Базова загальна середня	5
Початкова загальна	6
Не має початкової загальної	7
Неписьменний (на)	8
Дитина у віці до 6 років	9
Невідповідь	99

гр.9 "Коди наявності професійно-технічної освіти"	
Є	88
Немає	77
Дитина до 15 років	66
Невідповідь	99

гр.10 "Коди сімейного стану"	
Особі, які перебувають у шлюбі	1
Дорослі, які не перебувають у шлюбі (розлучені, вдівці, особи, які ніколи не перебували у шлюбі)	2
Особі до 18 років, які не перебувають у шлюбі	3
Невідповідь	9

гр.11 "Коди порядку запису дружини / чоловіка"	
Дружина, чоловік	порядковий номер запису
Особі відсутні більше 12 місяців	88
Невідповідь	99

гр.12, 13 "Коди порядку запису батьків для дітей у віці до 18 років"	
Мати, батько	порядковий номер запису
Не має батька /матері	77
Особі відсутні більше 12 місяців	88
Невідповідь	99

Продовження додатку Д

ОБСТЕЖЕННЯ СІЛЬСЬКОГОСПОДАРСЬКОЇ ДІЯЛЬНОСТІ НАСЕЛЕННЯ В СІЛЬСЬКІЙ МІСЦЕВОСТІ

3

РОЗДІЛ II. ХАРАКТЕРИСТИКА ДОМОГОСПОДАРСТВА

II.A ЗЕМЛЯ

1. Вкажіть, будь ласка, всю землю, якою користуються Ваше домогосподарство (включаючи орендовану, але за мінусом званої в оренду)

Кол	ВСЬОГО	У тому числі:			
		під житловим будинком, господарськими спорудами та парцям	земельні частки (пай), виділені в натурі	взято в оренду	решта
1	10	03	04	05	09
Площа, гектарів (з двома знаками після коми)	3				
Кількість ділянок	2				

2. Чи знаходиться земля, якою користуються Ваше домогосподарство, в єдиному масиві?

(потрібно позначити ✓)

Так 1 Вкажіть, будь ласка, відстань до найвіддаленішої ділянки км (з одним знаком після коми)

Ні 2

3. Як Ви оцінюєте рівень якості ґрунтів землі, якою користуються Ваше домогосподарство?

(дають узагальнену оцінку всіх ділянок домогосподарства)

(потрібно позначити ✓)

Роблять лише одну позначку

Хороша якість ґрунту 1

Середня якість ґрунту 2

Погана якість ґрунту 3

4. Яка площа сільськогосподарських угідь, якими користуються Ваше домогосподарство?

К о л	Рілля	Багаторічні насадження				Сіножа-ті, пасо-вища	Перело-ги	ВСЬОГО
		всього	сади	виногра-дники	ягідники			
1	01	02	03	04	05	06	07	10
Гектарів (з двома знаками після коми)	2							

Синожики, Фасівка, у літній період

5. Скільки у Вашому домогосподарстві є плодкових дерев, кущів, насаджень?

	Опи-сання ви-паро-вання	К о л	Всього		У т.ч. у плодо-носно-му віці	Опи-сання ви-паро-вання	К о л	Всього		У т.ч. у плодо-носно-му віці
			1	2				3	1	
Черешня	дерес	162				Сунця, полуниця	кв. метри	181		
Вишня	дерес	163				Маліна, ожина	кв. метри	182		
Слива	дерес	164								
Абрикос	дерес	165								
Яблука	дерес	171				Смородина всяка	кущі	183		
Груша	дерес	172				Агрус	кущі	184		
Горіх	дерес	176								
° волоський горіх	дерес	446								
						Виноград	кущі	190		

Продовження додатку Д

4 ОБСТЕЖЕННЯ СІЛЬСЬКОГОСПОДАРСЬКОЇ ДІЯЛЬНОСТІ НАСЕЛЕННЯ В СІЛЬСЬКОЇ МІСЦЕВОСТІ

6. Чи є у Вашому домогосподарстві теплиця або парник? (потрібне позначте ✓)

Так 1 Ні 2 До питання 8

7. Для вирощування яких культур Ви її/їого використовуєте?

	Код		Кв. метрів		Код		Кв. метрів		Код		Кв. метрів	
	1	2			1	2			1	2		
Овочі – всього	125			зелена цибуля	501			інші овочі	504			
○ огірки	126			редиска	502			Квіти зрізані і квіткова розсада	192			
○ помідори	127			капуста	503			Розсада	193			
Загальна площа теплиці / парника											001	

8. Під посів яких культур Ви використали землю для одержання урожаю в поточному році?

	Код		Площа, гектарів (з двома знаками після коми)		Код		Площа, гектарів (з двома знаками після коми)		Код		Площа, м. метрів
	1	2			1	2			1	2	
Зернові та зернобобові культури – всього	101			Технічні культури – всього	110			Овочі відкритого ґрунту – всього	130		
○ пшениця озима	401			○ цукровий буряк	111			○ огірки	131		
○ пшениця яра	402			○ соняшник	112			○ помідори	132		
○ ячмінь озимий	403			○ тютюн	114			○ капуста	133		
○ ячмінь ярий	404							○ кабачки	134		
○ жито озиме	406							○ перець солодий і гіркий	135		
○ жито яре	407							○ баклажани	136		
○ кукурудза на зерно	104			Баштани продовольчі культури	150			○ гарбузи столові	137		
				○ кавуни	432			○ цибуля на ріпку	417		
				○ дині	433			○ цибуля на перо	418		
				Кормові культури – всього	155			○ часник	139		
				○ коренеплоди	156			○ буряк столовий	140		
				○ сіяні трави	157			○ морква столова	141		
				Овочі відкритого ґрунту – всього	130			Квіти зрізані і квіткова розсада	920		
Картопля	120										

В С Ъ О Г О

100

гектарів (з двома знаками після коми)

Заповнює фахівець з інтерв'ювання

9. Яка площа ріллі у Вашому домогосподарстві не була засіяна для одержання урожаю в поточному році?

 гектарів

(з двома знаками після коми)

10. Чи дасте землю, що належить Вашому домогосподарству, в оренду? (потрібне позначте ✓)

Так 1

Вкажіть, будь ласка, площу даної в оренду землі

гектарів

(з двома знаками після коми)

Ні 2

Продовження додатку Д

ОБСТЕЖЕННЯ СІЛЬСЬКОГОСПОДАРСЬКОЇ ДІЯЛЬНОСТІ НАСЕЛЕННЯ В СІЛЬСЬКІЙ МІСЦЕВОСТІ

5

П.Б ХУДОБА ТА ПТИЦЯ

1. Чи утримується у Вашому домогосподарстві (потрібне позначте ✓):

	а) худоба	Так	1	Ні	2	До підрозділу П.В
	б) птиця	Так	1	Ні	2	

2. Яка саме худоба та птиця утримується у Вашому домогосподарстві?

	Код			Код			Код	
	1	2		1	2		1	2
Велика рогата худоба – всього	201		Вівці	215		Птиці всього віку – всього	240	
° корови	202		у т.ч. висцематки і ярики від 1 року і старші	216		° хури та шівні	241	
° бугаї-плідняки	203		Кози	220		у т.ч. хури дорослі	242	
° телиці від 1 до 2 років	204		у т.ч. козематки від 1 року і старші	221		° качки	245	
° телята до 1 року	206		Кони	225		° гуси	250	
° телиці від 2 років і старші	207		° кобили від 3 років і старші	227		° індички	255	
° інші вікові групи	209		° лоша́та до 1 року	226		° інша птиця	260	
Свині – всього	210		Кролі	230		Інші тварини (вказати яка):		
° свиноматки від 9 місяців і старші	211		у т.ч. кролематки	231				
° свинки ремонтні від 4 місяців і старші	212		Хутрові звірі	234				
° поросля до 2 місяців	213		° нутрії	235				
° інші вікові групи	214		° інші види хутрових звірів	239				
			Бджоли, сім'ї	270				

П.В ГОСПОДАРСЬКІ СПОРУДИ І ПРИМІЩЕННЯ, ОБЛАДНАННЯ І ТЕХНІКА

1. Чи є у володінні /користуванні Вашого домогосподарства які-небудь господарські споруди або приміщення? (потрібне позначте ✓)

Так <input style="width: 30px;" type="text" value="1"/>			
Ні <input style="width: 30px;" type="text" value="2"/>	Які саме і яка їхня площа?	К о д	Загальна площа спо- руд /приміщень, кв.метрів
		1	2

2. Яким чином, зазвичай, оброблюється земля у Вашому домогосподарстві? (потрібне позначте ✓)																									
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">° Вручну</td> <td style="width: 30px; text-align: center;">1</td> <td style="width: 30px; text-align: center;">01</td> </tr> <tr> <td style="padding: 2px;">° Конями /Волами</td> <td style="text-align: center;">1</td> <td style="text-align: center;">02</td> </tr> <tr> <td style="padding: 2px;">° Трактором</td> <td style="text-align: center;">1</td> <td style="text-align: center;">03</td> </tr> </table>	° Вручну	1	01	° Конями /Волами	1	02	° Трактором	1	03	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">° Для зберігання урожаю</td> <td style="width: 30px; text-align: center;">01</td> <td style="width: 30px; text-align: center;">01</td> </tr> <tr> <td style="padding: 2px;">° Для зберігання кормів</td> <td style="text-align: center;">02</td> <td style="text-align: center;">02</td> </tr> <tr> <td style="padding: 2px;">° Для утримання худоби та птиці</td> <td style="text-align: center;">03</td> <td style="text-align: center;">03</td> </tr> <tr> <td style="padding: 2px;">° Для утримання техніки та ремієнту</td> <td style="text-align: center;">04</td> <td style="text-align: center;">04</td> </tr> <tr> <td style="padding: 2px;">° Для комплексного використання у вище зазначених цілях</td> <td style="text-align: center;">05</td> <td style="text-align: center;">05</td> </tr> </table>	° Для зберігання урожаю	01	01	° Для зберігання кормів	02	02	° Для утримання худоби та птиці	03	03	° Для утримання техніки та ремієнту	04	04	° Для комплексного використання у вище зазначених цілях	05	05
° Вручну	1	01																							
° Конями /Волами	1	02																							
° Трактором	1	03																							
° Для зберігання урожаю	01	01																							
° Для зберігання кормів	02	02																							
° Для утримання худоби та птиці	03	03																							
° Для утримання техніки та ремієнту	04	04																							
° Для комплексного використання у вище зазначених цілях	05	05																							

Продовження додатку Д

6 ОБСТЕЖЕННЯ СІЛЬСЬКОГОСПОДАРСЬКОЇ ДІЯЛЬНОСТІ НАСЕЛЕННЯ В СІЛЬСЬКІЙ МІСЦЕВОСТІ

3. Чи є у Вашому домогосподарстві обладнання /техніка?

Якщо "ТАК", вкажіть які саме та їхню кількість

	Код	Кількість, штук		Код	Кількість, штук		Код	Кількість, штук
	1	2		1	2		1	2
<input type="checkbox"/> Плуг	01		<input type="checkbox"/> Крупорушка	08				
<input type="checkbox"/> Сівалка	02		<input type="checkbox"/> Автомобіль вантажний	09				
<input type="checkbox"/> Борона	03		Інше (вказати)					
<input type="checkbox"/> Культиватор	04							
<input type="checkbox"/> Трактор	05							
<input type="checkbox"/> Комбайн	06							
<input type="checkbox"/> Сепаратор	07							

РОЗДІЛ III. ДОДАТКОВІ ПИТАННЯ

1. Чи використовуєте Ви у веденні свого домогосподарства:

(потрібне позначте ✓)

	Код	Регулярно	За потребою	Ніколи
	1	2	3	4
<input type="checkbox"/> Мінеральні добрива	01			
<input type="checkbox"/> Органічні добрива	02			
<input type="checkbox"/> Засоби захисту рослин (вкл. пестициди)	03			
<input type="checkbox"/> Безпflugний основний обробіток землі (плоскорізами)	04			
<input type="checkbox"/> Районовані сорти сільськогосподарських рослин	05			
<input type="checkbox"/> Вапнування ґрунтів	06			
<input type="checkbox"/> Рекомендовану сівозміну	07			
<input type="checkbox"/> Іригацію земель (зрошення, осушення)	08			
<input type="checkbox"/> Племіньних плідників для запліднення тварин	09			
<input type="checkbox"/> Штучне запліднення тварин	10			
<input type="checkbox"/> Ветеринарні перевірки	11			
<input type="checkbox"/> Санітарну обробку тваринницьких приміщень	12			
<input type="checkbox"/> Санітарний контроль якості молока	13			

2. Чи залучає до сільськогосподарських робіт у Вашому домогосподарстві кого-небудь зі сторони (крім родичів)?

(потрібне позначте ✓)

Так 1 → Як часто? (потрібне позначте ✓)

Ні 2

<input type="checkbox"/> Постійно	<input type="checkbox"/>	01
<input type="checkbox"/> Для виконання сезонних робіт	<input type="checkbox"/>	02
<input type="checkbox"/> Разово	<input type="checkbox"/>	03

ФОРМА № 01-СТН

БАЗОВЕ ІНТЕРВ'Ю

ЗМІСТ

Передмова.....	3
Розділ 1. СТАТИСТИЧНІ СПОСТЕРЕЖЕННЯ ЯК МЕТОД ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ.....	5
Статистичне спостереження – перший етап статистичного дослідження.....	5
Джерела статистичних даних, вимоги, яким вони мають відповідати	9
Програмно-методологічні питання статистичного спостереження.	
Помилки спостереження.....	17
Організаційні питання статистичного спостереження	25
<i>Вправи та запитання для самоконтролю</i>	<i>29</i>
<i>Рекомендована література</i>	<i>33</i>
Розділ 2. ФОРМИ, ВИДИ ТА СПОСОБИ СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ.....	34
Організаційні форми статистичного спостереження: сутність, сучасні завдання і перспективи розвитку	34
Види статистичних спостереження за рівнем охоплення та часом реєстрації даних.....	39
Способи збирання та реєстрації статистичних даних.....	41
Сучасна практика застосування комбінованих видів і способів статистичного спостереження.....	42
<i>Вправи та запитання для самоконтролю</i>	<i>50</i>
<i>Рекомендована література</i>	<i>50</i>
Розділ 3. ПЕРЕПИСИ ТА СТАТИСТИЧНІ РЕЄСТРИ: СУТНІСТЬ, ЕВОЛЮЦІЯ, СУЧАСНА ОРГАНІЗАЦІЯ.....	53
Переписи як спеціально організована форма статистичного спостереження	53
Особливості сучасних переписів: міжнародний і вітчизня- ний досвід	59
Перепис населення: етапи проведення, програмно- методологічні та організаційні питання.....	66
Статистичні реєстри та формування статистичних банків даних	72

<i>Вправи та запитання для самоконтролю</i>	81
<i>Рекомендована література</i>	83

Розділ 4. СТАТИСТИЧНІ МОНІТОРИНГИ..... 84

Статистичний моніторинг як неперервне спостереження	84
Первинні та сучасні форми моніторингу	89
Економічний та фінансовий моніторинг.....	92
Соціальний моніторинг: соціологічні і статистичні обстеження.....	103
<i>Вправи та запитання для самоконтролю</i>	117
<i>Рекомендована література</i>	119

Розділ 5. ВИБІРКОВІ ОБСТЕЖЕННЯ: ЗАВДАННЯ І СПОСОБИ ОРГАНІЗАЦІЇ..... 120

Сутність, переваги і сфери використання вибірових обстежень .	120
Основа вибірки, типи основ	128
Способи формування вибірових сукупностей.....	135
Застосування не випадкового (спрямованого) добору.....	146
<i>Вправи та запитання для самоконтролю</i>	151
<i>Рекомендована література</i>	153

Розділ 6. ТОЧНІСТЬ ВИБІРКОВИХ ОЦІНОК..... 154

Функції і властивості вибірових оцінок	154
Стандартна похибка як міра точності вибірових оцінок	159
Дизайн вибірки та дизайн-ефект.....	168
Визначення мінімально достатнього обсягу вибірки	173
<i>Вправи та запитання для самоконтролю</i>	180
<i>Рекомендована література</i>	183

Розділ 7. ФОРМУВАННЯ СТРАТИФІКОВАНОЇ ВИБІРКИ..... 184

Методологічні принципи стратифікації.....	184
Способи розміщення (алокції) стратифікованої вибірки	188
Стратифікація у практиці державних вибірових обстежень	197
<i>Вправи та запитання для самоконтролю</i>	204
<i>Рекомендована література</i>	208

Роздл 8. БАГАТОСТУПЕНЕВІ ВИБІРКОВІ ОБСТЕЖЕННЯ	209
Схеми багатоступеневого (багатостадійного) добору	209
Оптимальне співвідношення обсягів вибірки на різних ступенях добору	216
Організація вибіркових обстежень населення (домогосподарств).....	221
<i>Вправи та питання для самоконтролю</i>	230
<i>Рекомендована література</i>	232
Роздл 9. ПОВТОРНІ СТАТИСТИЧНІ ОБСТЕЖЕННЯ І ПРИНЦИПИ РОТАЦІЇ	233
Схеми ротації одиниць вибірки	233
Оцінювання результатів ротаційної вибірки і забезпечення їх надійності.....	240
Особливості панельних обстежень.....	247
<i>Вправи та запитання для самоконтролю</i>	254
<i>Рекомендована література</i>	256
Роздл 10. ПОШИРЕННЯ РЕЗУЛЬТАТІВ ВИБІРКОВОГО ОБСТЕЖЕННЯ НА ГЕНЕРАЛЬНУ СУКУПНІСТЬ	257
Методичні засади поширення результатів вибірки на генеральну сукупність.....	257
Формування системи статистичних ваг для вибірок складного дизайну	264
Забезпечення надійності результатів обстеження на основі непрямих методів оцінювання.....	270
<i>Вправи та запитання для самоконтролю</i>	281
<i>Рекомендована література</i>	284
Список літературних джерел	285
Додатки А–В	288
Додаток Г — переписний лист	296
Додаток Д — запитальник	301

Навчальне видання

ЄРІНА Антоніна Михайлівна
ПАЛЪЯН Зінаїда Оганесівна

**СТАТИСТИЧНІ СПОСТЕРЕЖЕННЯ:
ПЕРЕПИСИ, МОНІТОРИНГИ,
ВИБІРКОВІ ОБСТЕЖЕННЯ**

Навчальний посібник

Редактор *В. Гломозда*
Художник обкладинки *Т. Зябліцева*
Коректор *В. Македон*
Верстка *О. Руденко*

Формат 60x84^{1/16}. Ум. друк. арк. 00,00. Наклад 300 пр.
Гарнітура Times. Папір офсет. Друк офсетний.
Підписано до друку 14.08.19.

Видавець і виготовлювач
Видавничо-поліграфічний центр «Київський університет»
01601, м. Київ, б-р Т. Шевченка, 14, кім. 43

Тел. (38044) 239 32 22; (38044) 239 31 72; тел./факс (38044) 239 31 28

e-mail: ypc@univ.kiev.ua

<http://ypc.univ.kiev.ua>

Свідоцтво суб'єкта видавничої справи ДК № 1103 від 31.10.02