

# ПСИХОЛОГІЧНИЙ СПЕЦПРАКТИКУМ ЗІ СПЕЦІАЛІЗАЦІЇ

ІУ КУРС, ГРУПИ П-41-42, 45

ВИКЛАДАЧІ: СТ.ВИКЛ. САВУЛЯК В.О., ДОЦ. РУДЮК О.В., ДОЦ. РОМАНЮК В.Л.

## ЗМІСТОВИЙ МОДУЛЬ 1. ПЛАНУВАННЯ ТА ОРГАНІЗАЦІЯ СОЦІАЛЬНО-ПСИХОЛОГІЧНИХ ДОСЛІДЖЕНЬ

### ТЕМА 1. Програма соціально-психологічного дослідження.

**Програма вивчення об'єкту.** У межах програми вирішується весь комплекс теоретичних і практичних задач дослідження (виробляються гіпотези, здійснюється операціоналізація понять, визначається метод і методики, можливість використання тих чи інших процедур і технічних прийомів; формується репрезентативна вибірка).

Програма уточнюється і конкретизується після проведення пілотажу розроблених методик та «зондажу» об'єкту і ситуації дослідження. Пробно-пошукове дослідження: 1) виявляє особливості всього процесу збору та обробки інформації; 2) дозволяє перевірити оптимальність місця, часу і ситуації проведення дослідження; 3) робить більш чітким припущення щодо структури об'єкту, причин тих чи інших процесів; 4) дозволяє уточнити деталі анкет, тестів, схему експерименту, організаційні форми дослідження, характер і об'єкти вибірки основного дослідження.

#### *Етапи здійснення програми:*

- 1) збір даних ("польовий" етап)
- 2) аналіз (представлення матеріалу у вигляді, придатному для аналізу, зіставлення, порівняння; статистичний аналіз)
- 3) встановлення надійності даних
- 4) змістовна інтерпретація результатів (звіт)
- 5) висновки і практичні рекомендації
- 6) впровадження
- 7) аналіз ефективності впровадження

Розрізняють:

- стратегічний план – включає теоретичні задачі (об'єкт, гіпотеза тощо);
- робочий план – розписані основні етапи, виділені за організаційним принципом і урахуванням хронологічної послідовності; вказуються строки і матеріальні витрати.

### ТЕМА 2. Суцільні та вибіркові дослідження. Види вибірок.

#### *Проблеми вибірки в соціальній психології*

**Основні поняття.** **Генеральна сукупність** дослідження (основний масив, популяція, вихідна множина, універсум) - об'єкт дослідження, що складається з певної досить великої кількості (N) елементів, кожен з яких є потенційною одиницею дослідження. Передбачається, що розподіл досліджуваної змінної, виявлений при вивченні деякої, можливо малої частини цих одиниць (n), тобто **вибірки**, буде характерним і для об'єкту, генеральної сукупності в цілому.

**Відбір** (вибір, визначення, ідентифікація) одиниць дослідження - спеціальна процедура цілеспрямованого виділення з генеральної сукупності більш-менш однозначно визначених елементів (вибірки) з подальшим обмеженням основного вивчення об'єкту на даній підмножині одиниць.

**Репрезентативність** - основна якісна (у широкому значенні) і кількісна (у вузькому значенні) характеристика вибірки. Іноді, хоча це не зовсім коректно, відповідно говорять і про репрезентативність методу відбору. Жодна вибірка не може бути абсолютно репрезентативною щодо генеральної сукупності за однією, а тим більше за комплексом ознак (аналогічно тому, як моделювання якогось об'єкту не передбачає створення абсолютно тотожного йому дублікату). З абсолютною репрезентативністю вибірка може представляти тільки саму себе.

Ступінь репрезентативності - це завжди **похибка** ( $\mu$ ), виправлення, необхідне для подання властивостей генеральної сукупності через вибірку.

Знаходження необхідного обсягу та структури вибірки для представлення певної генеральної сукупності з похибкою, що не перевищує наперед заданого значення, чи розрахунок величини похибки для заданих вибіркової й генеральної сукупностей становлять пряме завдання репрезентативності.

У той же час із деякою умовністю можна стверджувати, що будь-які відомості репрезентативні стосовно якоїсь сукупності явищ, що вивчені одиниці є частиною якогось більш об'ємного «надоб'єкту», тобто мають свою генеральну сукупність. Точніше, результати будь-якого професійно

проведеного дослідження мають репрезентативність різного ступеня стосовно декількох «надоб'єктів». Наприклад, у книзі «Людина та її робота» ми знайдемо міркування авторів про те, якою мірою отримані дані репрезентативні стосовно різних контингентів молоді. Дослідницька проблема при вирішенні такого зворотного завдання репрезентативності складається в чіткому зазначенні (виходячи з характеристик вибірки) відповідного їй «надоб'єкту» і його меж із урахуванням необхідного рівня точності.

Це завдання (із наперед заданою, задовольняючою дослідника похибкою) вирішується, зокрема, при використанні матеріалів, отриманих на повністю або частково «стихійних» вибірках. Наприклад, відповіді на запитання про те, яку сукупність (контингент) представляють: а) листи споживачів засобів масової інформації (при вивченні їхньої аудиторії); б) добровільні учасники соціально-психологічних експериментів; в) доступні респонденти в масовому опитуванні тощо.

Необґрунтовано розширене тлумачення результатів досліджень зустрічається частіше через ігнорування автором саме зворотного завдання репрезентативності. Причини цьому можуть бути різні, але найчастіше справа - у недостатнім ознайомленні дослідника із проблемами вибірки. Необґрунтовані висновки можуть з'являтися і в інтерпретаціях результатів дослідження окремими споживачами, замовниками соціально-психологічної продукції.

Зворотне завдання репрезентативності також вирішують, або, принаймні, повинні вирішувати, дослідники, що звертаються до вторинного аналізу через зіставлення даних, узятих з різних досліджень, присвячених одній проблемі. Принципи узагальнень і екстраполяції на ширшу спільність (генеральну сукупність) визначаються тут сукупною чисельністю одиниць спостереження, представленістю в такій вибірці основних типів одиниць, що становлять цю спільність, і припустимим рівнем похибки (з її імовірнісною оцінкою).

Це ж зворотне завдання репрезентативності стоїть перед дослідником, вибірка якого визначається якоюсь частковою фактологічною змінною, зафіксованою на момент збору інформації (нею може бути, наприклад, певна форма поведінки: у сфері дозвілля, праці, спілкування й ін.). Адже однією з головних цілей такого дослідження, як правило, є ідентифікація об'єкта наукового пошуку, тобто докладний соціально-психологічний опис усієї категорії осіб, що виконують дану дію в даних умовах. Тут вирішення зворотного завдання репрезентативності впритул наближається до вирішення питання щодо валідності, достовірності даних (методики) подібного соціально-психологічного дослідження.

Зі сказаного ясно, що проблема виявлення генеральної сукупності є специфічною для соціально-психологічних досліджень, у яких частіше, ніж, наприклад, у соціології застосовуються трудомісткі методи спостереження, глибокого інтерв'ювання, експерименту, лонгітюдного дослідження тощо, що спричиняють: а) відносно малу кількість одиниць при відборі; б) тяжіння соціальних психологів до доступної вибірки; в) гостроту проблеми валідності даних, одержаних у дослідженнях, особливо експериментальних.

У такій ситуації дослідник, що розробляє теоретичну проблему, природно, часто виявляється не в змозі забезпечити достатню відповідність свого дослідження вимогам чітко розрахованої репрезентативності. Відносна ж репрезентативність (у широкому значенні цього поняття) результатів може оцінюватися вже не через обчислення відповідної похибки, а:

- через проведення повторних досліджень на близьких за основними змінними чи тих самих (за умови достатньої їхньої стабільності) об'єктах (багатофазна вибірка),
- через виявлення прогностичних можливостей отриманих результатів,
- через змістовне зіставлення з результатами інших досліджень цієї чи близької проблеми, які проводилися на інших об'єктах із застосуванням іншого інструментарію,
- за допомогою змістовного співвіднесення отриманих результатів як зі сталими теоретичними положеннями, так і обов'язково з конкуруючими гіпотезами,
- шляхом застосування інших непрямих методів оцінки рівня репрезентативності.

Майже всі перераховані способи зазвичай розглядаються як засоби валідації результатів, але водночас вони є обґрунтуваннями вибірки. Однак у соціально-психологічних дослідженнях часто наявні можливості й для її статистичного обґрунтування. Таке обґрунтування особливо необхідно в прикладних дослідженнях, однією з основних особливостей яких часто є наявність чітко окресленої генеральної сукупності. Відхід же від принципів розрахунку репрезентативності у випадках, коли висновки дослідження поширюється на інші об'єкти, повинен завжди чітко зазначатися.

Важливо пам'ятати різні значення терміну «репрезентативність». Не цілком строге, широке тлумачення припускає лише принципову загальну оцінку подання цілого, класу через його частину. Наприклад, у цьому значенні вживається даний термін у роботах, що критикують невиправдано широку інтерпретацію результатів лабораторного експериментування на студентах-психологах.

У вузькому значенні слова під репрезентативністю розуміється кількісна оцінка ступеню

представленості чітко визначеної генеральної сукупності через конкретну вибірку. Тобто оцінка репрезентативності, у даної вузькому значенні терміну, це - імовірна оцінка розбіжностей між статистичними характеристиками вибірки й генеральної сукупності. Зі сказаного вище про специфіку підходу до проблеми вибірки в соціально-психологічному дослідженні зрозуміло, що й у ньому поняття «репрезентативність» може застосовуватися як у широкому, так і у вузькому значенні. Природнім наслідком цього є необхідність чіткого зазначення характеру його використання в кожному конкретному випадку.

Із репрезентативністю пов'язане уявлення про **адекватність** (у вузькому значенні) вибірки, однак вони не збігаються, хоча інколи їх не достатньо розрізняють. Вибірка є адекватною, якщо вона є досить великою, щоби бути впевненим у сталості її характеристик. Іншими словами, адекватною є така вибірка, розмір якої дозволяє при аналізі основної зібраної інформації одержати досить великі угруповання одиниць дослідження за сполученням окремих значень розглянутих ознак (не менше 15-20 одиниць в угрупованні). Це необхідно задля того, щоби будь-які можливі, але не істотні в даному дослідженні індивідуальні відхилення всередині цих угруповань взаємно погасались у ході зважування показників усередині них. Таким чином, поняття адекватності вибірки відображає лише один аспект її репрезентативності, а саме той, що відноситься до оцінки частини випадкових похибок вибірки.

### **Види відбору. Випадкові вибірки**

Для характеристики основних методів відбору варто зазначити можливі критерії їх класифікації:

➤ розрізнення апостеріорної та апріорної оцінок репрезентативності вибірки безпосередньо пов'язане із розрізненням відповідних можливих варіантів її створення, залежно від того, на якому етапі дослідження воно проведене. Якщо вибірка проектувалась до польового етапу і первинна інформація збиралася на її основі, то йдеться про **апріорний** варіант застосування вибірки. Якщо ж вибірка проектується для коригування стихійної доступної або неправильно здійсненої вибірки (через велику кількість або порушень інструкції інтерв'юером, респондентами, або відмов відповідей, або ж просто недоступних одиниць), тобто - для проведення "вибірки з вибірки", тоді в наявності **апостеріорний** варіант;

➤ **головний критерій** класифікації видів вибірки ґрунтується на розрізненні **випадкового** й **невипадкового** відбору. Ступінь випадковості у відборі - це континуальна характеристика: від строго випадкової вибірки через різновиди, в яких відбір проводиться не цілком випадково, до методів цілеспрямованого відбору, у яких імовірність потрапляння різних одиниць у вибірку регулюється багатьма доцільними обмеженнями. Іншими словами, підмножини випадкових і невідповідних вибірок істотно перекриваються. До того ж процедура конкретного відбору визначається ще й іншими критеріями. Тому, наприклад, віднесення якогось виду вибірки до категорії «випадкової» може бути досить умовним.

Таким чином, необхідно розмежувати: 1) **чисто (або власне) випадкову** й 2) **модифіковані випадкові** вибірки. Останні підрозділяються залежно від двох критеріїв. Відповідно до першого, розрізняють **районовані** і **нерайоновані** вибірки, а виходячи із другого критерію, виділяються **одноступінчасті** й **багатоступінчасті** вибірки. У свою чергу, районовані вибірки розрізняються на підставі районування (список, перелік тощо). Вона може бути **природньою**, тобто існуючою до дослідження, або спеціально **створеною** для нього. З вибірок із природньою основою районування в особливу категорію виділяються так звані територіальні вибірки, які дозволяють практично врахувати в дослідженні (як для підвищення репрезентативності даних, так і для економії засобів) територіальне розміщення досліджуваних одиниць. Однак природність основи в таких вибірках може бути відносною, оскільки необхідний план територіального розміщення одиниць, у якому враховані спеціальні інтереси даного дослідження, може бути складений **спеціально** для даного випадку.

Основні критерії для розрізнення різновидів невідповідної вибірки пов'язані з ролями тих, хто формує вибірку (дослідника, інтерв'юера тощо), і тих, кого вибирають (респондента, досліджуваного тощо), у прийнятті рішень про відбір;

● особливе місце займає критерій розрізнення **однофазних** і **багатофазних** вибірок. Він полягає у відсутності чи наявності часового інтервалу між вибірковими обстеженнями на одному об'єкті;

● конкретні види вибірки також нерідко розрізняють за критеріями, швидше пов'язаними із класифікацією самих досліджень, у яких вони застосовуються (тобто - за методом дослідження, його об'єктом, метою тощо). Наприклад, говорять про виробничу, підліткову або експериментальну вибірку. Подібні розмежування не зовсім коректні, однак у нестроному, «робочому» вживанні вони можуть бути корисними.

Типологію видів відбору природньо проводити в порядку, що відповідає головним критеріям розрізнення. При цьому необхідно також врахувати, що загальне удосконалення видів відбору проходило в напрямку відмови від строгої реалізації принципу випадковості на користь різних модифікацій процедур відбору (відповідно особливостям складних об'єктів, специфіці застосовуваних методів, економічним й організаційним обмеженням досліджень). Таким чином, раціоналізація видів відбору призвела: а) до скорочення витрат на дослідження великих об'єктів (насамперед, за рахунок скорочення числа одиниць спостереження без втрат у репрезентативності результатів); б) до певної «індивідуалізації» складних вибірок відповідно оригінальним особливостям того чи іншого дослідження; в) до зростання ролі фахівця з вибірки на всіх етапах сучасного соціального дослідження.

Однак теоретичною основою, а часто і складовим елементом складних сучасних вибірок є *випадкова ймовірнісна вибірка*. Деякі автори вважають навіть усі інші види вибірки не цілком науковими. З авторитетом випадкової вибірки пов'язане настільки широке поширення формул і рекомендацій щодо неї, що в деяких дослідженнях їхнє некритичне застосування призводить до неадекватних результатів. Грубі помилки можуть бути пов'язані, зокрема, з неправильним розумінням принципу випадковості. Як відомо, він полягає у вимозі забезпечити рівну ймовірність потрапляння у вибірку кожній одиниці досліджуваного об'єкту і реалізується в тім, що зі списку генеральної сукупності (з єдиною нумерацією одиниць) із використанням надійного засобу *рандомізації* (таблиць випадкових чисел, жеребкування тощо) відбирається необхідне число одиниць спостереження.

Однак зустрічається й повсякденне розуміння випадковості як ненавмисності відбору. Наприклад, неправильно називати випадковою вибірку (при дослідженні всього колективу підприємства), складену з робітників, яких інтерв'юер зустрів під час збору матеріалу на заводській території. Очевидно, такий відбір точніше назвати псевдовипадковим, оскільки у вибірці будуть слабо або зовсім не представлені категорії працівників, «жорстко» закріплених на робочому місці, а також працівники, типові маршрути внутрішньозаводського пересування яких або пролягають частиною території, не охопленої інтерв'юером, або припускають використання внутрішньозаводського транспорту тощо. Також метафоричним можна вважати називання «випадковим» відбору, що проводить журналіст, опитуючи перехожих на жвавому перехресті, з метою представити думку всіх жителів даного мікрорайону. У ледаря, що тиняється без діла, набагато більше шансів стати його респондентом, ніж у жителів, що працюють у цей час, або – людей похилого віку чи дітей, зона пересування яких обмежена двором. Подібні зсуви майже завжди трапляються при нестрогому дотриманні випадковості (не говорячи про зсуви, викликані до того ж суб'єктивністю безпосереднього відбору).

Власне випадкову вибірку застосовують порівняно рідко (обсяг генеральної сукупності не повинен перевищувати 800-1000 одиниць), тому що, з одного боку, її можна в більшості випадків помітно покращити, виходячи з економічних критеріїв, а з іншого боку, у багатьох ситуаціях її важко або навіть неможливо використати, наприклад, у силу відсутності (і практичної неможливості існування) єдиного списку генеральної сукупності.

*Механічний* (або систематичний) відбір є найпростішим за своєю процедурою і тому, можливо, найбільш розповсюджений у соціальних дослідженнях. Хоча він, як і чисто випадковий, передбачає наявність переліку одиниць, однак, при його здійсненні немає необхідності в об'єднанні їх у загальний список зі спільною нумерацією і застосуванні строгих засобів рандомізації. Все це дозволяє наблизити даний метод відбору до умов польового дослідження з використанням наявної «природньої» основи для відбору одиниць. При механічному відборі, виходячи з тієї долі, яку вибірка становить від генеральної сукупності, вибирається довжина (інтервал) кроку в  $k$  одиниць і у вибірку сукупність потрапляє кожна  $k$  - та одиниця з переліку.

Коли основа відбору, як це часто буває, складається відповідно змінам деякої характеристики у генеральній сукупності, наприклад, відтворює якийсь аспект її структури, механічний відбір забезпечує пропорційне представлення окремих значень цієї характеристики у вибірці (що наближує її до пропорційної *стратифікованої*). У разі зв'язку даної характеристики із досліджуваною це поліпшує результати механічної вибірки порівняно з випадковою.

Механічний відбір є модифікацією випадкового відбору і до того ж – найбільш наближеним до нього. Тому *формули для розрахунку похибок вибірових показників і необхідного обсягу (для середньої і для частки) при механічному відборі застосовуються ті самі, що і при випадковому*.

До числа специфічних недоліків механічного відбору відносять можливість зсуву вибірки тоді, коли в основі відбору (у списку) генеральна сукупність розташована в порядку, що відображає періодичні коливання деякої характеристики, пов'язаної із досліджуваною ознакою. Цей зсув буде особливо сильним, коли таке коливання виявиться кратним крокові (інтервалу) механічного відбору. Для запобігання такого роду систематичних помилок рекомендується, по-перше, ретельно перевірити

основу в даному відношенні, по-друге, у ході відбору можна неодноразово, через якесь число кроків, посувати початкову одиницю на деяке випадкове число одиниць, по-третє, можна вибрати на початку списку кілька вихідних одиниць і здійснювати відбір через однакові чи різні інтервали ніби паралельно.

### **ТЕМА 3. Обчислення похибок для вибірок різного виду**

#### **Помилки вибірки й непрямі методи їх оцінки**

Оскільки здійснення вибіркового дослідження завжди супроводжується похибками, завданням дослідника є не повне їх усунення, а мінімізація, контроль і урахування при інтерпретації результатів. Похибки, що виникають при застосуванні вибіркового методу можуть мати різні причини. Справа в тому, що відповідність методу відбору структурі досліджуваних ознак в об'єкті і цілям дослідження, ретельність у дотриманні процедур конструювання вибірки і при безпосередньому відборі одиниць у польових умовах, а також інші моменти, що впливають на репрезентативність вибірових даних, не досягаються лише завдяки застосуванню формальних засобів, а залежать перш за все від компетентності і добросовісності виконавців.

Похибки бувають двох видів:

- *похибки спостереження* (реєстрація, вимірювання), неминучі в будь-якій дослідницькій процедурі
- *похибки репрезентативності* (власне похибки вибірки), можливі тільки у вибіркового дослідженні

Помилки спостереження, зумовлені перш за все методом збору матеріалу, певною мірою співвідносяться з помилками репрезентативності. Здійснення відбору ставить дослідника перед дилемою кількість (збільшення обсягу вибірки, що зменшує помилку репрезентативності) – якість (більш ретельна робота з одиницею вибірки, що знижує помилку спостереження). Вирішення проблеми – у збалансованому підході.

Серед помилок репрезентативності розрізняють помилки систематичні і випадкові. *Систематичні* призводять до так званого зміщення даних дослідження. Це зміщення викликається достатньо стабільним впливом якогось неконтрольованого фактора на відбір конкретних одиниць у вибірку. Практичною передумовою дії такого фактора є принципова помилка у конструюванні вибірки чи систематичні відхилення від дотримання правил відбору. Наприклад: заміна, недобір, псевдовипадковий відбір досліджуваних одиниць зі свідомо прийнятою чи неусвідомленою тенденцією щодо специфічно гомогенної частини вибірки. При цьому зміщення буває особливо значним у разі прояву тенденції щодо переваг чи уникання певних одиниць, котрі мають властивості, неявно пов'язані з досліджуваною ознакою.

Особливий різновид систематичної помилки репрезентативності пов'язаний із тим, що власне факт відбору, тобто факт безпосередньої участі в дослідженні, може призводити до значних змін у характеристиках одиниць, включених у вибірку. Така помилка тісно пов'язана з відповідною похибкою вимірювання. У якості неконтрольованих факторів можуть виступати: 1) недосконалість методики збору інформації (наприклад, відсутність психологічного контакту в інтерв'ю, що зачіпає інтимні сторони життя респондентів, але не передбачає гарантію анонімності); 2) обставини, що супроводжують застосування методики (наприклад, плітки, тенденційна думка у досліджуваній групі щодо подальшого використання отриманих у дослідженні результатів); 3) збурювання, що привноситься у досліджувані характеристики одиниць самим фактом присутності дослідника на момент збору інформації чи його стійкими особистісними якостями. Оскільки для соціально-психологічних досліджень характерний тісний контакт дослідника з об'єктом, імовірність систематичних помилок, викликаних останньою обставиною, істотно зростає. Цим, зокрема, пояснюється місце даної проблеми в сучасній методологічній літературі із соціальної психології. Прикладом прийому, спрямованого на мінімізацію подібного зсуву, є методична вимога - пояснювати обстежуваним причини включення їх у вибірку, тобто випадковий або невідповідний характер відбору.

Практика застосування вибіркового методу знає чимало випадків навмисних систематичних помилок репрезентативності. Це визначається головним чином тим, що результати масових опитувань можуть бути важливим важелем впливу на суспільну думку в політику й економіці. Занепокоєння громадськості у зв'язку з таким станом справ призвело до вимог одержання додаткової інформації від організаторів будь-якого опитування про джерело фінансування, основні методи інтерв'ювання, обсяг вибірки й методи відбору, відсоток ухилень від відповідей, формулювання питань тощо. Систематичні помилки розглядаються чи не всіма дослідниками в якості найбільшої небезпеки в межах вирішення проблем репрезентативності. Загальний принцип переважання якості над кількістю при цьому зводиться до правила: увага до того, які одиниці вибрати (тобто уникання систематичних помилок),

важливіше прагнення до збільшення вибірки (тобто до зменшення випадкових помилок).

*Випадкові помилки* вважаються менш небезпечними, тому що вони в меншому чи більшому ступені мають тенденцію взаємно погашати одна одну: одні перебільшують справжнє значення ознаки, інші применшують. Ці помилки, тією мірою, в якій вони мають вигляд нормального розподілу й можуть бути віднесені до статистично великих чисел, піддаються точному вимірюванню на основі розрахунків із використанням відомих елементів апарату математичної статистики.

Поряд із поділом помилок репрезентативності на систематичні і випадкові, навмисні й ненавмисні, виділяються також фактичні й теоретичні похибки.

*Фактична похибка* репрезентативності розраховується як різниця між значеннями деякої характеристики генеральної сукупності, відомої дослідникові з якихось інших джерел (наприклад, деякої її середньої -  $\bar{X}$ ), і значенням цієї характеристики, отриманим на вибірці в даному дослідженні (відповідної середньої -  $\bar{x}$ ). При цьому процентний рівень репрезентативності, похибка вибірки за цією загальною характеристикою становить:

$$\Delta\% = \frac{\bar{X} - \bar{x} \cdot 100}{\bar{x}}$$

Більшість дослідників прийнятною вважає похибку, що не перевищує **5%**. Величина фактичної похибки являє собою результат сукупної дії систематичних і випадкових помилок на репрезентативність вибірки за обраною для порівняння загальною характеристикою. Однак у конкретному дослідженні, особливо в соціально-психологічному, зазвичай націленому на «ментальні», настановні змінні, дійсна оцінка репрезентативності, підрахованої цим способом, набагато складніше.

Така оцінка передбачає розгляд ступеню тісноти зв'язку між обраною для зіставлення загальною характеристикою й основною змінною дослідження. Справа в тому, що порівняння вибірки з усім об'єктом зазвичай можна провести лише за характеристиками об'єкту, наявними в офіційному обліку (винятки становлять випадки, коли вибіркове дослідження проводиться на тім же об'єкті безпосередньо після суцільного, у якому розроблялася близька тема). Таким чином репрезентативність вибірки за загальносоціальними, кваліфікаційними, освітніми чи іншими характеристиками є лише непрямим підтвердженням репрезентативності за досліджуваною соціально-психологічною характеристикою.

Знання фактичних похибок репрезентативності вибірки за одними змінними тією мірою можна переносити на інші змінні, в якій дослідник має аргументи щодо тісного зв'язку між цими змінними. Такі аргументи можна отримати як у власному дослідженні, так і з даних близьких за темою досліджень інших об'єктів.

Чим більш проблематичним є зв'язок досліджуваної характеристики з тією, за якою розраховано фактичну похибку вибірки, тим більш непрямим є даний спосіб визначення рівня репрезентативності вибірки. На практиці застосовується декілька власне непрямих методів перевірки репрезентативності, для здійснення яких дані про розподіл характеристик у досліджуваній генеральній сукупності взагалі не потрібні.

Найбільш розповсюджений із цих методів ґрунтується на порівнянні результатів досліджень, проведених одним або різними авторами на одному й тому самому чи на подібних об'єктах за спільною або різними методиками. Однак побічно результати таких зіставлень можуть інтерпретуватися і як уточнення ступеню репрезентативності вибіркових досліджень.

Даний метод припускає велику кількість варіантів порівняльного аналізу для різних поєднань вихідних умов. Можливі як порівняння незалежних, важко порівнюваних досліджень, так і зіставлення результатів досліджень, близьких за темою й проведених одним автором на сукупностях, які перетинаються чи навіть збігаються. Останній варіант особливо ймовірний на об'єктах, з якими проводиться тривала дослідницька робота.

Особливою модифікацією даного методу непрямої перевірки репрезентативності (і достовірності) інформації є повторні, перевірочні дослідження - їх називають *реплікаціями*.

Проблеми репрезентативності, пов'язані з різними варіантами порівняльного аналізу досліджень, є фактично проблемами повторного аналізу. Простої констатації збігу даних або їхньої суперечності недостатньо, адже можлива загальна систематична помилка. У кожному разі необхідно змістовне й методичне зіставлення досліджень у цілому.

На ідеї порівняльного аналізу ґрунтується метод незалежних підвбірок. Цей метод застосовується для непрямої оцінки репрезентативності як на польовому етапі збору основної інформації, так і на етапі її первинного аналізу. Загальний обсяг вибірки випадковим способом ділиться на декілька ( $r$ ) підвбірок, середні показники яких ( $\bar{x}_i$ ) порівнюються із середнім усієї вибірки ( $\bar{x}$ )

$$\mu_i = \sqrt{\frac{\sum \bar{x} - \bar{x}_i^2}{r(r-1)}}$$

Прийнятні значення таких похибок ( $\mu_i$ ) для кожної підвибірки відносно всієї вибірки (за досліджуваними показниками) розглядаються як непрямі свідчення прийнятності відхилень  $\mu_i$  і показників вибірки ( $\bar{x}$ ) відносно відповідних показників об'єкту ( $\bar{X}$ ).

Варіантом методу незалежних підвбірок є також прийом, коли похибка вибірки А, поділеної на підвбірки (наприклад, k, l, m), визначається як:

$$\mu_A^2 = \frac{\bar{x}_{A-k} - \bar{x}_k^2 + \bar{x}_{A-l} - \bar{x}_l^2 + \bar{x}_{A-m} - \bar{x}_m^2}{r(r-1)}$$

де  $\bar{x}_{A-k}$  – середня для всіх елементів вибірки за винятком елементів підвбірки k, тобто для елементів підвбірок l і m.

У соціологічній практиці при застосуванням методу незалежних підвбірок зазвичай використовується від 2 до 10 підвбірок, у соціально-психологічній - 2-3. Зрозуміло, що метод незалежних підвбірок не виявляє систематичних помилок, обумовлених недоліками конструювання всієї вибірки. Проте, він особливо корисний, коли обчислення теоретичних похибок і зіставлення з показниками генеральної сукупності здійснити важко, що досить типово для соціально-психологічних досліджень. Даний метод застосовується також для пошуку евристичним шляхом необхідного обсягу вибірки. Коли ж підвбірки сформовані на підставі того, хто з дослідників безпосередньо збирав інформацію, цей метод може бути використаний для перевірки якості їхньої роботи, насамперед їх "індивідуальних" зсувів (за умови контролю інших можливих факторів таких зсувів) або вирішення інших методичних питань.

Усі перераховані вище способи непрямой оцінки репрезентативності - це методи більш-менш точного визначення фактичної помилки вже отриманих дослідником або наявних незалежно від нього вибірових показників, тобто - *апостеріорні* методи оцінки вибірки. Однак частіше трапляється і є більш природною зворотна ситуація, коли дослідник повинен оцінити передбачувану репрезентативність вибірових даних, які він тільки планує зібрати. Іншими словами, він повинен *апріорно* оцінити або розрахувати *теоретичну* (передбачувану, очікувану) помилку репрезентативності. Точне обчислення цієї категорії похибок доступне лише тоді, коли передбачається строго випадковий відбір на всіх етапах вибірки. Таким чином, апарат обчислення випадкових похибок тим менше придатний для оцінки раціоналізованих вибірок, чим більшою мірою при їхньому створенні порушений принцип випадковості відбору. Теоретичні помилки таких не цілком випадкових вибірок повинні оцінюватися евристично якісними методами.

Виділяють три основних взаємозалежних фактори, що визначають величину похибок репрезентативності:

- ◆ *структурованість - однорідність* генеральної сукупності за досліджуваними змінними. Абсолютно однорідну сукупність елементів можна представити через зазначення їхньої кількості та значень змінних будь-якого з них. З іншого боку, сукупність, у якій кожен елемент є у досліджуваному аспекті неповторною індивідуальністю, передбачає дослідження кожного з них, тобто суцільне. Таким чином, навіть приблизне знання амплітуди коливань досліджуваних ознак надзвичайно важливе для раціонального конструювання вибірки. Важливо пам'ятати про наступне: одним із факторів, що „визначають” амплітуду коливань (суб'єктивну – для дослідника) досліджуваної ознаки, є методичні засоби її вияву у дослідженні. Чим вони досконаліші, тим більшою є їхня диференційна спроможність стосовно градацій досліджуваної ознаки. Наприклад, за однакових інших умов, для відповідей на відкриті запитання опитувальників характерний зазвичай набагато більший розмах, ніж для відповідей на ті самі, але закриті запитання. Варіації даних при виявленні переваг, виборів тощо значною мірою визначається кількістю альтернатив (експлікованих чи можливих, існування котрих неявно припускається);

- ◆ ступінь кількісної представленості (*частки*) об'єкту у вибірці та *кількість одиниць* у ній. Для великих генеральних сукупностей у цьому реалізується «закон великих чисел»;

- ◆ застосування адекватного *методу відбору одиниць* із вихідної сукупності.

Дотримання правил побудови випадкової вибірки дозволяє застосувати для розрахунку середньої (або стандартної) теоретичної похибки вибіркової середньої ( $\mu$ ) основну формулу

$$\text{вибіркового методу: } \mu = \sqrt{\frac{\sigma^2}{n}} \text{ або } \mu = \frac{\sigma}{\sqrt{n}} \quad (1)$$

де  $\sigma^2$  - дисперсія ознаки в генеральній сукупності,

$n$  - обсяг вибірки.

Дисперсію ознаки можна оцінити приблизно, якщо відоме значення розмаху для генеральної



сукупності. Задля цього використовується співвідношення зв'язку середнього квадратичного відхилення з розмахом  $\sigma \approx (x_{\max} - x_{\min}) : V$ , чинним при припущенні нормального характеру розподілу. У цій формулі  $(x_{\max} - x_{\min})$  – варіаційний розмах генеральної сукупності, а  $V$  – величина, котра залежить від обсягу вибірки, див. табл.:

|                  |     |     |     |     |     |            |
|------------------|-----|-----|-----|-----|-----|------------|
| Обсяг вибірки, n | 5   | 10  | 20  | 30  | 50  | $\geq 100$ |
| Значення V       | 2,3 | 3,1 | 3,7 | 4,1 | 4,5 | 5,0        |

Якщо досліджувана ознака є не кількісною (такою, наприклад, як вік, розряд, соціометричний статус, інтенсивність установки або рівень розвитку групи, виражені балами шкали тощо), а якісною (стать, професія, стиль лідерства, знак установки, рівень розвитку групи, виражений загальною характеристикою) і визначається часткою одиниць (p), що мають такі якості, то

$$\sigma^2 = p(1 - p), \mu = \sqrt{\frac{p(1-p)}{n}} \quad (2)$$

Очевидно, що при дуже малих значеннях  $p \approx \sqrt{\frac{p}{n}}$  (3)

Відбір залежно від того, чи може одиниця сукупності, що вже раз потрапила у вибірку, бути знову в неї включеною, буває *повторним* і *безповторним*. У переважній більшості соціальних досліджень, насамперед з етичних міркувань, застосовується безповторна вибірка, тобто перший же номер зі списку (основи), що потрапив у вибірку, відразу з нього виключається, і наступний вибір фактично здійснюється вже не з  $N$ , а з  $N-1$  одиниць, що залишилися в генеральній сукупності. Тому при безповторній випадковій вибірці величина похибки множиться на відповідний поправочний коефіцієнт, що характеризує частину, на яку скоротилася вихідна сукупність у процесі відбору, тобто

$$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)} \quad (4) \quad \text{або} \quad \mu = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)} \quad (5)$$

Цей коефіцієнт зменшує величину середньої похибки вибіркового показника тією мірою, наскільки великою є частка вибірки в генеральній сукупності.

Мета знаходження значення  $\mu$  полягає в тім, що воно допомагає із заданою ймовірністю визначити похибку вибірки –  $\Delta$ , тобто інтервал, у якому знаходиться шукане середнє генеральної сукупності (при припущенні щодо відсутності помилок зсуву):

$$\bar{x} - \mu < X < \bar{x} + \mu \quad (6)$$

У всіх наведених формулах  $\sigma^2$  є характеристикою ступеню коливань досліджуваної ознаки в генеральній сукупності. Однак у переважній більшості випадків дослідник не може наперед знати дисперсію досліджуваної ознаки для визначення необхідних показників вибірки; оскільки її знаходження входить до числа цілей дослідження. Для виходу із цього протиріччя при розрахунках фактичної похибки вибірки практично використовується відповідний вибіркового показник ( $\sigma$  або  $p$ ), а для визначення теоретичної похибки, якщо йдеться про якісні ознаки, доводиться орієнтуватися на максимально можливу величину розсіювання  $p(1 - p) = 0,25$ , якщо ж – про кількісні, то оцінка дисперсії досліджуваної ознаки повинна проводитися (у випадку відсутності даних про неї з минулих досліджень) за попередньо отриманим на об'єкті результатам пробного обстеження або експертним оцінкам.

Оцінка похибки вибірки може бути лише ймовірнісною, оскільки очевидним є істотний вплив на неї рівня вимог щодо точності (надійності) кінцевих вибіркового показників, котрий ми самі задаємо. Адже відомо з практики: чим точнішим має бути результат виміру, тим важче його досягти, і, навпаки, чим менші вимоги, тим легше вони можуть бути задоволені. Згідно з теоретико-ймовірнісною оцінкою впливу необхідної надійності отриманого показника, наведена формула (6) відповідає ймовірності  $P = 0,68$  потрапляння шуканого показника об'єкту (наприклад, середньої) у даний інтервал. Тобто, 0,68 – це той ступінь ймовірності, з якої можна очікувати, що середня генеральної сукупності  $X$  виявиться в межах  $\bar{x} - \mu < X < \bar{x} + \mu$  (тобто  $\Delta = \mu$ ). Відповідно, 0,32 – ймовірність того, що ця середня виявиться поза цим інтервалом.

Жорсткість вимог до надійності результату призводить до розширення цього інтервалу. Так, для більш надійної його оцінки зручно задати рівень ймовірності  $P = 0,95$ . Тоді інтервал можливих значень середньої об'єкту стає вдвічі ширшим, тобто  $\bar{x} - 2\mu < X < \bar{x} + 2\mu$ . Однак усе ще залишається ймовірність  $P = 0,05$ , що шукана характеристика вибіркової сукупності розташована поза ним. Припустимо, потрібно одержати ще надійніший результат – з ймовірністю  $P = 0,997$ . Для цього первісний інтервал треба розширити вже в 3 рази:  $\bar{x} - 3\mu < X < \bar{x} + 3\mu$  (тобто  $\Delta = 3\mu$ ). Таким чином, ймовірнісний характер вибірки визначається додатковим, до  $\mu$ , коефіцієнтом довіри  $t$  ( $\Delta = t \cdot \mu$ ), певні



значення якого відповідають ймовірностям:

|          |        |
|----------|--------|
| t = 1    | 0.683; |
| t = 1,96 | 0,950  |
| t = 2    | 0.954; |
| t = 2,58 | 0,990  |
| t = 3    | 0.997; |
| t = 4    | 0.999. |

При застосуванні цього апарату важливо враховувати, що він адекватний тією мірою, у якій розподіл досліджуваних ознак близький до нормального, особливо у випадках, коли обсяг вибірки відносно невеликий.

Вибіркову сукупність, що складається з порівняно невеликої кількості одиниць (20...30), називають *малою вибіркою*, характеристики якої також можна поширювати на генеральну сукупність. При малих вибірках дисперсію обчислюють з урахуванням кількості *ступенів вільності варіації* – кількості одиниць, здатних змінюватися після того, як для всіх одиниць вибірки було визначено їх загальну характеристику. Середня похибка малої вибірки:

$$\mu_{\text{М.В.}} = \sqrt{\frac{s_{\text{М.В.}}^2}{n-1}} = \frac{s_{\text{М.В.}}}{\sqrt{n-1}}; \text{ де } s_{\text{М.В.}} = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}; n-1 - \text{число ступенів вільності варіації,}$$

яке вказує на кількість різних можливих значень варіантів з їх середньою арифметичною. Аби зв'язати похибку малої вибірки з граничною, враховують те, що в разі недостатньо великого обсягу вибірки стандартизована різниця між вибірковою і генеральною середньою має розподіл Стюдента, не нормальний. Значення t при невеликому обсязі вибірки наведено у спеціальних таблицях.

Дотепер йшлося про оцінку репрезентативності випадкової вибірки (про розрахунок її похибок) лише за кожною окремою із усього комплексу ознак, що зазвичай вивчається в дослідженнях. Природно, що значення похибок для різних ознак будуть істотно варіювати. Чи можливе обчислення інтегральної репрезентативності для всієї сукупності виявлених у дослідженні вибірових показників? Адже побудова вибірки у великому дослідженні завжди передбачає досягнення комплексу цілей, тому гостро відчувається необхідність її нехай різнобічної, але цілісної оцінки. Загальноприйнятих варіантів рішення поки немає.

### ***Розв'язати задачі; сформулювати висновки.***

1. Для вивчення думки 680 студентів факультету щодо проведення певних заходів методом випадкового безповторного відбору опитано 60 осіб; з них 40 підтримали план заходів. З імовірністю 0.954 визначте межі для частки тих, хто підтримав заходи, у генеральній сукупності.

2. Середній стаж роботи працівників підприємства становить 10.8 років при середньому квадратичному відхиленні 3,3 року. Визначте з імовірністю 0,997 граничну похибку вибірки та інтервал, в якому знаходиться середній стаж роботи.

Варіантом випадкової вибірки, що наближається за своїм характером до районованої, є *проста серійна вибірка*. Вона є особливо зручною, коли природно основу відбору поділено на рівні (чи майже рівні) за обсягом групи - *серії* (наприклад: бригади, класи, сім'ї тощо). Тоді проста серійна вибірка формується з випадково або механічно відібраних серій із наступним суцільним вивченням одиниць у них. Природно, що застосування серійного відбору тим доцільніше, чим більше дослідник упевнений у близькості внутрішньосерійної дисперсії досліджуваної ознаки до її дисперсії по об'єкту в цілому. У цьому випадку серійна вибірка не тільки доречна, але репрезентативність її різко зростає за рахунок мінімізації межсерійної дисперсії. Головною перевагою серійної вибірки є організаційна: вивчення декількох груп одиниць, кожна з яких територіально локалізована, значно легше, ніж вивчення такого ж числа одиниць, але просторово розкиданих.

Очевидно, що в разі серійної вибірки, яка передбачає суцільне вивчення одиниць у відібраних серіях, похибка вибірки залежатиме не від числа обстежених одиниць, а від кількості відібраних серій. Формування серійної вибірки на основі випадкового або механічного відбору серій визначає використання відповідних формул для розрахунку теоретичних похибок (формули **4,5,7,8**). При цьому об'єкт і вибірка розглядаються як сукупності серій і, відповідно, дисперсійні показники – як *дисперсії серійних середніх*. Похибка вибірки залежатиме не від варіацій ознаки в сукупності, а від варіації серійних середніх, яка вимірюється міжсерійною (міжгруповою) дисперсією  $\delta^2$ , показниками: **S** – число серій у генеральній сукупності, **s** – число відібраних серій.

Гранична похибка серійної вибірки:

| Спосіб відбору | Визначення середньої | Визначення частки |
|----------------|----------------------|-------------------|
|----------------|----------------------|-------------------|

|              |   |   |
|--------------|---|---|
| Повторний    | $\Delta = t \sqrt{\frac{\sigma_x^2}{s}}$                              | $\Delta = t \sqrt{\frac{\sigma_p^2}{s}}$                              |
| Безповторний | $\Delta = t \sqrt{\frac{\sigma_x^2}{s} \left(1 - \frac{s}{S}\right)}$ | $\Delta = t \sqrt{\frac{\sigma_p^2}{s} \left(1 - \frac{s}{S}\right)}$ |

### Районовані (не цілком випадкові, раціоналізовані) вибірки

Принцип районування спирається на відому теорему про додавання дисперсій, з якої слідує, що загальна дисперсія  $\sigma^2 = \sigma_m^2 + \sigma_s^2$  де  $\sigma_m^2$ - міжгрупова, а  $\sigma_s^2$ - середня внутрішньогрупова дисперсія. Підвищення репрезентативності районованих вибірок досягається за рахунок зменшення обох доданків або одного з них.

Завдяки непряму або безпосередньому урахуванню в районованих вибірках дисперсії досліджуваної характеристики, вони завжди будуть за обсягом менше, ніж випадкова або механічна вибірки однакової з ними репрезентативності. Таким чином, при цих видах відбору особливо важливе значення для конструювання вибірки має апіорний опис об'єкту дослідження.

*Стратифікована* вибірка, що часто називають просто "районованою" або "типовою", є, у даній класифікації, одним із різновидів районованих вибірок. Сутність її зводиться до того, що процедуру відбору націлено на зменшення обох доданків формули (9), але особливо - внутрішньогрупової дисперсії. Об'єкт, представлений повним переліком, ділиться, як правило, на нерівні групи одиниць (страги, типові райони) так, щоби кожна страта була якомога більше однорідною, тобто містила в собі одиниці із найбільш близькими значеннями досліджуваної ознаки. Число страт при використанні однієї ознаки стратифікації може бути в інтервалі від 3 до 10. Ефект стратифікації краще позначається при вивченні кількісних ознак.

Стратифіковані вибірки у свою чергу можуть розрізнятися за: а) відбором з кожної страги кількості одиниць пропорційно чи непропорційно її обсягу; б) використанням при виділенні страт однієї або декількох ознак (*багатомірна* стратифікація); в) формуванням страги на основі евристичних міркувань або з використанням формальних методів; г) ступенем випадковості відбору всередині страт.

Найпростіший варіант стратифікованої вибірки передбачає використання для виділення страт однієї ознаки, формування страт на основі приблизних міркувань і застосування пропорційного випадкового відбору в стратах. Результатом цієї процедури відповідно буде *одномірна пропорційна стратифікована вибірка*.

Обсяг вибірки з окремої страги ( $n_i$  з  $N_i$ ) при пропорційному відборі визначається за формулою:  $n_i = n \frac{N_i}{N}$ . Формули розрахунку репрезентативності для простої стратифікованої вибірки ті ж, що і для випадкової (формули 4,5,7,8), з тією тільки зміною, що дисперсійні показники в них даються як *середні з дисперсій страт*.

Додаткова раціоналізація стратифікованої вибірки зазвичай передбачає, що поряд із членуванням генеральної сукупності на страги, останні ранжуються у відповідності з передбачуваною варіацією в них досліджуваної характеристики (величиною дисперсії). При проведенні відбору всередині страт розташування тієї чи іншої з них у цьому проранжованому ряду є підставою для непропорційного відбору з них, оскільки більш однорідні за досліджуваною ознакою страги можуть бути представлені у вибірці (без збільшення її помилки) меншим числом одиниць, ніж різномірні.

Можливі і змістовні, обумовлені специфікою дослідження підстави для непропорційної стратифікації. Наприклад, вибірка в дослідженні, представленому в книзі «Людина і її робота», була непропорційною через особливі кількісні відмінності досліджуваних страт та орієнтацію дослідників на якісний аналіз об'єкту.

Переваги точного розрахунку репрезентативності при пропорційній стратифікованій вибірці часто не можуть бути використані через суворість процедурних вимог, а саме: необхідність переліку всіх одиниць генеральної сукупності й строгої рандомізації відбору всередині страт. До того ж, повне дотримання правил випадковості призводить до того, що одиниці, включені у вибірку, виявляються дуже розпорошеними територією, на якій розміщений об'єкт, що значно ускладнює організаційний бік дослідження (збільшення транспортних і відрядних витрат, необхідного числа дослідників, загальних строків тощо).

Мінімізувати вплив цих негативних моментів дозволяє інший різновид районованого відбору - *багатоступінчаста гніздова вибірка*. Її суть зводиться до розрізнення в об'єкті проміжних і кінцевого об'єктів (і відповідно, ступенів відбору), що співвідносяться один з одним за принципом "матрьошки". Тобто, основний об'єкт на першому проміжному кроці відбору розглядається як сукупність деяких

великих проміжних одиниць, відібрані з них для другого ступеня розглядаються також як сукупність, але інших одиниць, більш дрібних, ніж на першому, зате - більших, ніж на третьому кроці й т.д. ... до останнього ступеню, одиницями якого будуть виступати властиво одиниці дослідження (наприклад, чотириступінчастий ланцюжок вибіркової сукупності з однієї промислової галузі: підприємства - цехи - бригади - робітники). У практиці соціальних досліджень нерідкі вибірки, що включають до п'яти ступенів. Ясно, що тут немає необхідності у загальному спискові основних одиниць (у наведеному прикладі - всі робітники цілої галузі); на кожному кроці складається "свій" список проміжних одиниць, з якого здійснюється випадковий відбір. Таким чином, частка вибірки в генеральній сукупності

$p = p_1 \cdot p_2 \cdot \dots \cdot p_k$ , де  $k$  - число ступенів відбору.

Об'єкт, що відбирається на кожному проміжному кроці в якості вихідної сукупності для наступного відбору, називається гніздом (кінцеві одиниці не є гніздами). У практиці соціальних досліджень використовуються найрізноманітніші типи гнізд: регіони й населені пункти, міські райони, поштові відділення, виборчі й лікарські дільниці, будинки, поїзди, вагони, установи, цехи, виробничі ділянки, студентські групи; сукупності осіб, що включені у картотеки лікарень і поліклінік або опинились на випадково визначених сторінках списку виборців, будинкової книги, класного журналу; гнізда за тимчасовим критерієм, наприклад, газети тільки за певні числа, дні тижня тощо.

Переваги раціоналізації вибірки при даному відборі не тільки організаційно-економічні. Вони пов'язані з основною особливістю гніздової вибірки, яка, на противагу стратифікованій, що особливо ефективна при мінімальній дисперсії всередині страт, найбільш гарантована від зсувів при максимальній внутрішньогніздовій кореляції. Із цією особливістю пов'язаний додатковий спосіб оцінки репрезентативності гніздової вибірки через коефіцієнт внутрішньогніздової кореляції. Чим він вище, тим більше негативний прояв «ефекту гнізда», і навпаки. Даний спосіб дозволяє уникати при гніздовому відборі істотних систематичних помилок. Збільшення при цьому загальної випадкової помилки вибірки за рахунок збільшення другого доданка у формулі (9) не відбувається, завдяки компенсації через зменшення міжгніздової дисперсії, тобто за рахунок першого доданка у формулі (9). Дана перевага особливо помітна при внесенні в гніздовий відбір елементів стратифікації.

На різних ступенях відбору, а часто й на всіх, вихідний "природний" список проміжних одиниць попередньо стратифікується. Відбір гнізд здійснюється зі страт, а отримана в такий спосіб вибірка є вже «гібридом» гніздового й стратифікованого методів відбору. Відбір гнізд із попередньо сформованих страт найкраще робити за допомогою процедур власне випадкової або механічної вибірок. У випадках, коли через організаційні обмеження число відібраних гнізд повинно бути мінімальним, їх бажано мати не менше двох із кожної страти. Зокрема, це дозволяє використати надалі, для додаткової непрямої перевірки репрезентативності, метод незалежних підвибірок.

Відбір гнізд, особливо на перших етапах, часто здійснюється також виділенням «типових одиниць», які представляють страти не на основі випадковості їхнього вибору, а відповідно до *експертної оцінки* за декількома показниками. Така оцінка іноді доповнюється формальними засобами (наприклад, розрахунок типових одиниць як «центральної у своєму таксоні»). Незважаючи на те, що відбір зі страт типових одиниць позбавляє дослідників можливості точно визначити випадкову похибку репрезентативності, він широко розповсюджений через його явну евристичну цінність.

При гніздовому відборі зі страт (у тому числі і при виділенні типових одиниць) важливо забезпечити пропорційність представництва страт у гніздах. Вона не тільки підвищує загальну репрезентативність даних, але й дозволяє, де це можливо, наприклад, для одноступінчастої гніздової (серійної) вибірки, розрахувати за зазначеними формулами випадкову (незміщену) похибку вибіркової середньої або частки, а також необхідну чисельність вибірки при вивченні середньої і частки.

Похибка середньої для багатоступінчастої (у  $K$  щаблів) гніздової вибірки складається з відповідних похибок на кожному щаблі:  $\mu = \sqrt{\mu_1^2 + \mu_2^2 + \dots + \mu_k^2}$ . Повний вид формул розрахунку такої похибки дається у відповідних посібниках. Для приблизної оцінки *двоступінчастої* вибірки використовується спрощений прийом, при якому її похибка, розрахована як для випадкової, множить на  $\sqrt{2}$  (тобто на 1,414...).

Похибки гніздової вибірки залежать також від розмаху величини гнізд на кожному етапі. Спеціальні операції, спрямовані на збереження принципу випадковості, при цьому полягають у забезпеченні пропорційності між імовірністю потрапляння того або іншого гнізда у вибірку і його величиною. Для цього, наприклад, можна найменше за розміром гніздо брати за одиницю, а всі інші гнізда характеризувати числом відповідних одиниць. На практиці через організаційні міркування зазвичай надають перевагу, за інших рівних умов, гніздам більшої величини.

### Розв'язати задачі; сформулювати висновки.

1. При перевірці відвідування студентами аудиторних занять було отримано такі результати:

| Курс             | I   | II   | III  | IV   | V    |
|------------------|-----|------|------|------|------|
| Обсяг вибірки    | 50  | 50   | 50   | 50   | 50   |
| Частка присутніх | 1,0 | 0,93 | 0,84 | 0,51 | 0,33 |

На підставі вибіркового даних вирахувати оцінку частки присутніх для генеральної сукупності ( $N=850$ ,  $P=0,954$ ).

2. Для виявлення розміру наданих кредитів державні та комерційні банки регіону здійснили 5%-у типову вибірку з добором одиниць пропорційно чисельності типових груп (в середині груп використовувався метод випадкового безповторного відбору). За наведеними результатами визначте з імовірністю 0,997 інтервал, в якому знаходиться середній розмір кредитів, наданих банками регіону.

| Типи банків | Кількість наданих кредитів | Середній розмір кредиту, млн.грн. | Середнє квадратичне відхилення |
|-------------|----------------------------|-----------------------------------|--------------------------------|
| Державні    | 160                        | 17,2                              | 3,8                            |
| Комерційні  | 300                        | 32,6                              | 6,2                            |

### ТЕМА 4. Розрахунки необхідного обсягу вибірки.

Поряд із визначенням теоретичних похибок випадкової вибірки дослідникові доводиться вирішувати завдання щодо теоретичного, апріорно необхідного обсягу ( $n$ ) вибірки, що задовільняв би наперед заданим вимогам щодо її майбутньої репрезентативності. Дане завдання визначається загальною ідеєю вибіркового дослідження - зберегти необхідний рівень точності результатів при можливо максимальному скороченні витрат на дослідження об'єкту.

Розрахунок необхідного обсягу одноступінчастої випадкової вибірки проводиться за формулами:

$$\begin{aligned} \text{для кількісних ознак} \quad n &= \frac{t^2 \cdot \sigma^2}{\Delta^2} && \text{при повторному відборі} \\ n &= \frac{t^2 \cdot \sigma^2 \cdot N}{\Delta^2 \cdot N + t^2 \cdot \sigma^2} && \text{при безповторному} \end{aligned} \quad (7)$$

$$\begin{aligned} \text{для якісних ознак} \quad n &= \frac{t^2 p(1-p)}{\Delta^2} && \text{при повторному відборі} \\ n &= \frac{t^2 p(1-p)N}{\Delta^2 N + t^2 p(1-p)} && \text{при безповторному} \end{aligned} \quad (8)$$

Таким чином, очевидно, що головними факторами, котрі визначають необхідний обсяг вибірки, будуть: дисперсія досліджуваної ознаки ( $\sigma^2$ ), обсяг генеральної сукупності ( $N$ ), вимоги до рівня точності результатів ( $\Delta$ ) та їх надійності ( $t$ ).

Слід зазначити, що явно помилковою є думка, що обсяг вибірки повинен становити деякий певний відсоток генеральної сукупності. Єдино правильним способом визначення цього обсягу для випадкової вибірки та її модифікацій є застосування відповідних формул розрахунку.

Із формул (7-8) видно, що головний шлях скорочення обсягу випадкової вибірки без втрат у репрезентативності матеріалу полягає в попередньому якомога точнішому визначенні ступеню коливань досліджуваної ознаки в генеральній сукупності. Тому цілком себе виправдує витрата зусиль на її визначення на підготовчій стадії дослідження: пошук відповідних вказівок у дослідницькій практиці, попередній зондаж об'єкту, експертні опитування тощо. Загалом, можна стверджувати, що сама по собі кількість вивчених у вибіркового дослідженні одиниць не є показником його масштабів, швидше, навпаки. Рівень професіоналізму дослідника більше проявляється в одержанні обґрунтовано надійних, цікавих висновків про об'єкт на порівняно малій кількості вивчених одиниць.

### Розв'язати задачі; сформулювати висновки.

1. На підприємстві, де працює 3 тис. робітників, необхідно встановити середній стаж роботи. Попередніми обстеженнями встановлено, що середнє квадратичне відхилення стажу роботи становить 5 років. З імовірністю 0,997 обчисліть необхідний обсяг механічної вибірки за умови, щоби похибка не перевищувала 1 рік.

2. В академії навчається 10 тис. студентів, з них 5 тис. – денної ф.н., 4 тис. – заочної ф.н., 1 тис. - вечірньої ф.н. Скільки студентів потрібно відібрати для визначення середнього балу за результатами екзаменаційної сесії із застосуванням типової вибірки з випадковим безповторним відбором всередині типових груп, щоби з імовірністю 0,954 похибка вибірки не перевищувала 0,5 бала. Дисперсія за даними попередніх досліджень дорівнює 0,23.

3. Серед робітників заводу, що має 200 бригад, проектується серійна вибірка для дослідження, однією з цілей якого є встановлення частки бригад (по заводу) з оптимальними показниками психологічного клімату. Яким повинен бути обсяг цієї вибірки, щоби, за можливою приблизною оцінкою, її гранична похибка не перевищувала 5%? За результатами недавно проведеного соціометричного дослідження можна припустити, що міжсерійна дисперсія частки 225.

## ТЕМА 5. Формування не випадкових вибірок.

### Невипадкові (спрямовані) вибірки. Багатофазний відбір

Невипадкові вибірки не передбачають дотримання принципу випадковості, а, отже, і рівноймовірності потрапляння кожної одиниці вихідної сукупності у вибірку. Тому репрезентативність цих вибірок може бути оцінена кількісно тільки за фактичними (але не за теоретичними) помилками або непрямими методами.

Основними причинами звернення до не випадковому відбору є: етичні принципи соціального дослідження (передбачається, що участь у ньому є суто добровільною), а часто й обмеженість засобів і часу, відпущених на дослідження. Звідси впливають і головні критерії для розрізнення конкретних форм не випадкової вибірки: а) стихійність її формування, тобто готовність дослідника включати у вибірку тих, хто бажає брати участь у дослідженні та охоче надає інформацію; б) цілеспрямованість відбору, тобто прагнення дослідника формувати вибірку з одиниць генеральної сукупності не стільки зручних, скільки з певними ознаками. Основна відмінність між цими діючими одночасно факторами полягає в тому, що включення у вибірку розглядається як результат рішення або досліджуваного, або дослідника. Розглянемо конкретні види не випадкової вибірки на перетині двох зазначених критеріїв, що припускають більше або менше обмеження свободи дій дослідника й досліджуваних.

| Вибірка формується ....<br>досліджуваними...   | ...дослідником як... |                   |
|--|----------------------|-------------------|
|  | ... доступна         | ...цілеспрямована |
| ... у мінімальному ступені                     | I                    | IV                |
| ...у максимальному ступені<br>(тобто стихійна) | II                   | III               |

(Нумерація різновидів не випадкової вибірки відповідає порядку їх подальшого розгляду.)

У різновиді I склад вибірки свідомо визначений дослідником заздалегідь як доступний, а в II - доступність одиниць виявляється тільки в ході збору інформації. До вибірки II часто застосовується визначення «акцидентальна». Прикладом тут, зокрема, є так званий «нюрнберзький метод», коли вибірка складається з колег, знайомих тощо. Сюди ж часто можуть бути віднесені власне *зондажні* вибірки в пробних дослідженнях із методичними чи ознайомлювальними цілями. До цього ж різновиду II відносяться доступні вибірки при вивченні інтимних сторін життя людей (або - до I, якщо відбиратися буде заздалегідь визначена категорія людей). Строго випадковий відбір у таких дослідженнях не придатний хоча би тому, що відсоток відмов при ньому перевищив би будь-яку норму.

Доступні вибірки I і II типів застосовуються також у випадках, коли деякий масив вивчається по частині його одиниць, доступних через те, що вони зафіксовані в спеціальному переліку (наприклад: представники певної категорії хворих, виділені із наявної картотеки лікувальної установи; неблагополучні підлітки - із зареєстрованих у дитячій кімнаті міліції; неблагополучні сім'ї, ідентифіковані за заявами про розлучення в суді або РАГСі тощо). Звичайно, екстраполяція отриманих у таких вибірках даних на всю досліджувану сукупність певною мірою умовна, чому й потрібно додаткове ретельне обґрунтування правомірності виділення в дослідженні того чи іншого «надоб'єкту».

Доступна вибірка найширше використовується в журналістиці, кінодокументалістиці тощо, тобто в сферах, де дотримання строгих вимог не настільки обов'язкове. Широко розповсюджений варіант доступної - стихійна (II) вибірка вивчається в ситуації, коли основна інформація отримана в результаті часткового повернення опитувальника, зверненого до всієї генеральної сукупності або до її репрезентативної частини (наприклад, сукупність повернутих анкет, опублікованих у періодичному виданні, або при іншому заочному, поштовому, опитуванні).



Специфічним прикладом цілеспрямованої вибірки з істотним впливом на відбір згоди респондентів (III) є вибір досліджуваної одиниці при монографічному обстеженні. Зазвичай мета такого обстеження полягає у виявленні ролі розповсюджених явищ, більш-менш інваріантних щодо соціально-демографічних або контекстуальних факторів.

До III типу не випадкових вибірок відносяться й цільові вибірки. Вони застосовуються: а) при відборі членів експериментальних і контрольних груп, котрі дали згоду на участь в експерименті і мають певні властивості; б) при «експертній вибірці», тобто при відборі експертів, заснованому як на їхній згоді ними бути, так і на наявності їхніх необхідних характеристик (об'єктивних - документованих і текстових, а також суб'єктивних - за самооцінками та оцінками інших осіб).

Основні типи цілеспрямованого відбору, що припускає мінімальний вплив досліджуваних на формування вибірки (IV), це - методи типових одиниць і квотований. *Метод типових одиниць* уже згадувався при розгляді питання про виділення типових гнізд. При необхідності згоди цих одиниць на включення у вибірку наближається до III категорії видів не випадкового відбору, зокрема, до монографічного методу. Головні попередні підстави для виділення репрезентативних типових одиниць, як правило, не формальні, а засновані на змістовному аналізі й експертних оцінках.

Принцип *квотованого відбору* зводиться до того, що дослідник, не покладаючись на механізм випадку, сам шукає потрібні йому одиниці сукупності, котрі мають певні поєднання характеристик. Квотований відбір є найбільш регламентованим зі всіх різновидів не випадкової вибірки, оскільки його проведення, як правило, припускає майже повний збіг вибірки з апріорною моделлю генеральної сукупності за обраними параметрами. Так само, як і в усіх інших аналогічних випадках, репрезентативність, а, отже, і ефективність квотованої вибірки в конкретному дослідженні визначається обґрунтованістю виділення даних параметрів як таких, що тісно корелюють із досліджуваними характеристиками. В іншому разі фактично вивчений у вибірці аспект об'єкту (генеральної сукупності) виявляється не релевантним його досліджуваному аспекту. Цим найчастіше й пояснюються серйозні помилки, що мали місце в історії застосування квотованого відбору.

Цілеспрямоване формування вибірки, а отже, і мінімальний її обсяг у поєднанні з певною свободою при відборі конкретних одиниць дослідження, є істотними організаційними перевагами квотованого методу, що обумовлюють ступінь його поширеності в соціальних дослідженнях. Зазвичай при квотованому відборі використовують 3-4 ключових ознаки, частіше - соціально-демографічних. Вони повинні бути максимально незалежними одна від одної, оскільки їхня взаємозалежність може звести нанівець зазначені переваги квотованої вибірки. При більшій кількості ознак відбір респондентів стає занадто трудомістким.

Залежно від того, до якого ступеня близькості вибірки до об'єкту прагне дослідник, розрізняють два основних варіанти квотованого відбору. Більш строгий варіант відрізняється від менш строгого тим, що вибіркова сукупність при ньому представляє генеральну не тільки за середнім значенням обраних характеристик, але й за часткою одиниць, котрі мають певне сполучення ознак. При цьому більш строгому варіанті число груп, що становлять вибірку й контролюються дослідником, стає загрозливо великим (якщо А, В, С - ключові ознаки при відборі, а А', В', С' - відповідне число значень (градацій) кожної з цих ознак, то загальне число таких груп у вибірці = А'·В'·С'). Крім того слід враховувати, що: а) співвідношення цих груп у вибірці повинно бути пропорційним їхньому співвідношенню в об'єкті; б) розмір кожної з них повинен задовільняти вимогам адекватності. Через це частіше застосовується інший, менш строгий варіант квотованої вибірки.

Основні умови ефективності квотованої вибірки: 1) ретельне попереднє вивчення об'єкту на основі «свіжих», не застарілих даних про нього; 2) контроль за роботою збирачів інформації, з метою уникання зсувів і реєстрації кількості відмов (особливо на останніх етапах збору); 3) включення, де це можливо, у квотовану вибірку елементів випадкової. Останнє може бути забезпечено, наприклад, тим, що інтерв'юер керується у квотованому відборі не тільки вказівками про необхідну кількість носіїв тих чи інших значень за ключовими ознаками, але також і випадково складеним списком осіб, із числа яких він повинен вибирати відповідних респондентів. Інший спосіб більшої «рандомізації» квотованої вибірки полягає у застосуванні її на останньому етапі багатоступінчастої випадкової вибірки.

На закінчення розглянемо особливий вид вибірки, критерій виділення якої не пов'язаний із характеристикою відбору як випадкового чи не випадкового. Це - *багатофазна* (ротаційна, стадіальна) вибірка. Вона має місце при кількаразовому вибірковому обстеженні якоїсь генеральної сукупності. Багатофазна вибірка відрізняється від багатоступінчастої тим, що у цьому випадку завжди зберігається одна й та сама одиниця відбору, а обстеження відібраних одиниць проводиться на кожній фазі.

Багатофазні вибірки відрізняються одна від одної в залежності: 1) від кількості фаз; 2) від



частки одиниць із вибірки першої фази, знову включених у вибірки другої та наступних фаз; 3) від дотримання принципу випадковості при відборі одиниць із вибірок попередніх фаз у наступні вибірки; 4) від величини часового інтервалу, що відокремлює фази одна від одної; 5) від співвідношення обсягів вибірок.

Найчастіше на практиці застосовуються двофазні вибірки. Залежно від того, чи є спільна частина у вибірок першої й другої фаз, виділяють наступні типи двофазної вибірки: (а) вибірки повністю збігаються, тобто дослідження є панельним; (б) вибірка другої фази становить частину від першої, тобто є, властиво, вибіркою з вибірки (частково панельне дослідження); (в) вибірки обох фаз збігаються лише частково, тобто лише перетинаються; (г) вибірки зовсім не збігаються (таке обстеження одного і того ж об'єкту називають повторним).

Повторні обстеження мають ряд переваг над панельними: 1) відпадає необхідність забезпечувати сталість складу вибірок; 2) зменшується число відмов; 3) знімається ефект впливу попереднього контакту з одиницями.

Двофазними є також дослідження, в яких на підготовчих етапах, з метою попереднього вивчення об'єкту (наприклад, одержання інформації про «вагу» страт) або відпрацювання методики, застосовується пробне (відповідно, зондажне або пілотажне) дослідження. Вибірка, застосовувана в ньому, значно менше вибірки основної фази і, як правило, є не випадковою. Обсяг вибірки в пробному обстеженні, на думку багатьох авторів, повинен становити не менше 5-10% обсягу основної вибірки (або не менше 100 одиниць). Окрім методичного використання результатів пробної вибірки, їх також можна, із деякими застереженнями, застосовувати для додаткового підтвердження матеріалів основного етапу. Для такого застосування особливо зручні пробні вибірки, репрезентативні хоча би за однією характеристикою, близькою до досліджуваної.

Головною проблемою при здійсненні панельних досліджень є труднощі, а часто й неможливість забезпечення однакового складу, і навіть досить близької структури, вибірки на різних фазах. Причинами цього є: природні зміни в складі об'єкту й у характеристиках одиниць, а також можливі відмови й недоступність деяких одиниць на другій фазі. Панельне дослідження важко здійснити також у випадках, коли при одержанні інформації на першій стадії зберігалася анонімність.

За допомогою багатофазної вибірки дослідниками вирішується широке коло теоретичних, методичних і організаційних завдань. Наприклад, типовий спосіб вирішення питання щодо сталості, відтворюваності отриманих даних (або, навпаки, їхньої зміни внаслідок експериментального впливу), а також щодо надійності методики, полягає у проведенні двофазного дослідження (зазвичай із коротким часовим інтервалом). Другий приклад пов'язаний із завданнями здешевлення, спрощення дослідження шляхом цілеспрямованого відбору одиниць, коли основний, розширений інструментарій застосовується на другій фазі до вибірки з вибірки першої фази, обстеженої за більш вузькою методикою.

Багатофазне дослідження, нерідко з великими інтервалами між фазами, передбачається й тоді, коли йдеться про багатозрізове або лонгитюдне дослідження. Як правило, тут потрібен один із варіантів (краще варіант *a*, ніж *b*) панельного дослідження.

*Взаємопроникна вибірка* потрібна тоді, коли із однієї генеральної сукупності незалежні дослідники одним і тим самим способом здійснюють кілька незалежних вибірок. Взаємопроникні вибірки надають незалежні одна від одної оцінки значень досліджуваної сукупності, і якщо результати різних вибірок близькі між собою, то такі результати дуже переконливі. До недоліків слід віднести порівняно великі видати на їх проведення. Похибки взаємопроникної вибірки визначаються за формулами типової пропорційної вибірки.