

## ЛЕКЦІЯ 2 ВИБІРКОВИЙ МЕТОД

1. Поняття генеральної сукупності і вибірки.
2. Види вибірок: випадкова, серійна, механічна, типова, комбінована.
3. Методи формування випадкових вибірок.

### 1. Поняття генеральної сукупності і вибірки.

Психолог-експериментатор у більшості випадків вивчає певну визначену вибірку людей, яка завжди відбирається із більшої за чисельністю групи.

Наприклад, дослідник припускає, що перегляд телепередач, що містять сцени насилля, підвищує рівень агресії у підлітків. Об'єктами-носіями властивостей в цьому випадку будуть всі підлітки. Але для перевірки даної гіпотези необхідно виміряти рівень агресії у всіх підлітків, що зробити неможливо. Тому при проведенні такого дослідження обмежуються лише невеликою групою представників відповідної сукупності людей.

Множина всіх об'єктів, які підлягають дослідженню, називають **генеральною сукупністю**, а підмножина випадково відібраних об'єктів із генеральної сукупності, називають **вибірковою сукупністю** або **вибіркою**.



Теоретично вважається, що об'єм генеральної сукупності не обмежений. Практично ж об'єм генеральної сукупності завжди обмежений і може бути різним в залежності від предмета спостереження і тієї задачі, яку ставить перед собою психолог.

Практично всі психологічні дослідження є вибірковими, а їх висновки розповсюджуються на генеральні сукупності.

**Вибіркою** називається будь-яка підгрупа елементів (досліджуваних, респондентів), виділених із генеральної сукупності для проведення експерименту. При цьому окремий індивід з вибірки, з яким працює психолог, називається респондентом (досліджуваним).

**Об'ємом сукупності** (вибіркової або генеральної) називають кількість об'єктів цієї сукупності. Об'єм вибірки (позначається літерою  $n$ ) може бути будь-яким, але не меншим, ніж 2 респондента.

**Гомогенною** або однорідною називається сукупність, всі характеристики якої притаманні кожному її елементу;

**Гетерогенною** або неоднорідною називається сукупність, характеристики якої зосереджені в окремих підмножинах елементів.

Змістом вибіркового методу є розв'язання таких задач:

- Яку треба зробити вибірку?

- Яким повинен бути її об'єм?
- Які допустимі межі відхилення характеристик генеральної сукупності, за яких вибірку можна вважати репрезентативною?

Вибірка в експериментальних дослідженнях складається з **експериментальної** групи – випробовуваних, що піддаються впливу незалежної змінної і **контрольної** групи – випробовуваних, що знаходяться в тих же самих умовах, за виключенням незалежної змінної. Контрольна група забезпечує точку відліку, з якою порівнюють результати експериментальної групи.

### **Залежні і незалежні вибірки**

При порівнянні двох (і більше) вибірок важливим параметром є їх залежність. Якщо можна встановити гомоморфну пару (тобто, коли одному випадку з вибірки X відповідає один і тільки один випадок з вибірки Y і навпаки) для кожного випадку в двох вибірках, такі вибірки називаються залежними.

Приклади залежних вибірок:

- пари близнюків,
- чоловіки і дружини і т. п.

У разі, якщо такий взаємозв'язок між вибірками відсутній, то ці вибірки вважаються незалежними, наприклад:

- чоловіки і жінки,
- психологи і математики.

Відповідно, залежні вибірки завжди мають однаковий обсяг, а обсяг незалежних може відрізнятися.

Визначення якості вибірки включає опис низки умов, які дають змогу оцінити *валідність*, *надійність* та *репрезентативність* вибіркової сукупності стосовно об'єкта дослідження.

Для того, щоб за вибіркою можна було досить впевнено судити про властивості генеральної сукупності, вибірка має бути репрезентативною.

**Репрезентативність** вибірки означає, що об'єкти вибірки досить добре переставляють **генеральну сукупність**. Це забезпечується випадковістю відбору, тобто всі об'єкти мають однакову ймовірність потрапити до вибірки.

Генеральна сукупність неоднорідна – вона має свою структуру, оскільки включає людей різних класів, професій, вікових параметрів, статі тощо.

Якщо вибірка репрезентативна – то виявлені на ній закономірності можна перенести на генеральну сукупність.



Існують різні ступені репрезентативності, і зрозуміло чому – жоден дослідник не в змозі сформувавши вибірку, яка б абсолютно точно відтворювала структуру генеральної сукупності. Це призвело б до того, що вибірка стала б рівною по обсягу самій генеральній сукупності. Тому дослідник мусить обирати – великі обсяги вибірки і дуже високу репрезентативність, або малі за обсягом вибірки і не дуже високу репрезентативність.

Існують загальні методи підвищення репрезентативності вибірки:

**1. Планомірний підбір досліджуваних із генеральної сукупності.** При використанні цього методу вивчають структуру генеральної сукупності та обирають для дослідження представників усіх виділених категорій. Цей метод неекономний і доволі громіздкий. Крім того, яким би точним не був аналіз генеральної сукупності, є ймовірність зробити помилку і не врахувати якісь категорії.

Дослідник резонно може припустити, що хлопці та дівчата різняться як за рівнем агресивності, так і за рівнем сприйняття телевізійних сцен насилля. Якщо дослідник планує узагальнити вплив телебачення на агресивність усіх підлітків, то, керуючись соціально-демографічними даними, він має забезпечити ідентичне генеральній сукупності співвідношення хлопців і дівчат у вибірці.

**2. Рандомізований підбір досліджуваних з генеральної сукупності.** Цей метод ґрунтується на уявленні про ймовірнісну характеристику розподілу ознак. Суть його полягає у підборі досліджуваних з допомогою генератора випадкових чисел (або таблиць). У цьому випадку економиться час, матеріальні ресурси і є висока ймовірність попадання у вибірку більшості категорій генеральної сукупності.

Вивчаючи агресивність підлітків, дослідник може випадковим чином зупинити свій вибір на трьох класах різних шкіл, а потім випадковим чином відібрати по 10 учнів із кожного класу. Якщо ж дослідник просить учня для дослідження запросити своїх друзів, він грубо порушує принцип випадковості відбору.

**Надійність та валідність.** Надійність та валідність — основні параметри визначення якості інформації, одержаної в результаті проведення емпіричного дослідження. Під **надійністю** розуміють певну гарантію, що одержаний результат правильно відбиває досліджувану дійсність. **Валідністю** (обґрунтованістю) інформації називається підтвердження (доказ), що досліджувалися (вимірювалися) саме ті явища, які психолог передбачав дослідити.

Отримані з допомогою статичного спостереження матеріали можуть бути відповідним чином оброблені і на основі обробки можна зробити певні висновки про досліджуваний процес чи явище. Однак, на висновок можуть вплинути **помилки і похибки**, які виникають при дослідженні.

**1. Помилки реєстрації** (виникають внаслідок неправильного запису фактів):

- *випадкові* (описки, недостатнє знання і т.д.),
- *систематичні*:
- *навмисні* (коли опитувані чи реєстратор навмисно повідомляють чи записують неправильні дані),
- *ненавмисні* (зумовлені систематичними випадковими причинами – несправність приладів, втома досліджуваних після робочого дня тощо).

**2. Помилки репрезентативності** (характерні лише для несучільного спостереження – виникають внаслідок того, що склад відібраної для дослідження вибірки не відтворює усієї сукупності досліджуваних).

Виявити помилки допомагає статистичний контроль (арифметичний та логічний).

## **2. Види вибірок: випадкова, серійна, механічна, типова, комбінована.**

В статистиці розрізняють *малу* ( $n < 30$ ), *середню* ( $30 < n < 100$ ) і *велику вибірку* ( $n > 100$ ). Малі вибірки використовуються при статистичному контролі відомих властивостей вже вивчених сукупностей. Великі вибірки використовуються для установки невідомих властивостей і параметрів сукупності.

За способом відбору одиниць для дослідження розрізняють такі різновиди вибіркового спостереження:

- **власне випадкова вибірка** (повторна та неповторна) - якщо елементи відбираються випадковим чином. Так як більшість методів математичної статистики ґрунтується на понятті випадкової вибірки, тобто природно вибірка повинна бути випадковою;
- **механічна** - коли вся сукупність ділиться на стільки частин, скільки одиниць планується в вибірці і потім з кожної частини відбирається один елемент;
- **типова** - сукупність ділиться на гомогенні частини, і з кожної здійснюється випадкова вибірка;
- **серійна** (гніздова) - сукупність ділять на велике число різновеликих серій, потім роблять вибірку однієї будь-якої серії;
- **комбінована** - поєднуються розглянуті види відбору, на різних етапах.

**Власне випадкова вибірка.** За такого способу відбору всі одиниці генеральної сукупності мають однакову можливість потрапити в досліджувану групу. Відбір одиниць проводять жеребкуванням.

Приклад власне випадкової безповторної вибірки – тираж виграшів лотереї: усі номери випущених лотерейних білетів кладуть в урну; ретельно перемішують і витягують із неї один за одним наперед задану кількість виграшних номерів.

Якщо, наприклад, потрібно визначити вік 1000 студентів, то на картки однакового розміру записують фактичний вік, кладуть картки в урну; витягнувши з неї одну картку, записують номер і знову кладуть картку в урну, ретельно перемішують картки, знову виймають одну й так до заданої чисельності вибірки. Це приклад *повторної* вибірки.

**Механічна вибірка.** Це різновид власне випадкової вибірки, коли всі одиниці генеральної сукупності розміщують у певному порядку:

- за алфавітом;
- часом реалізації (виробництва);
- розміщенням у просторі та ін.

Потім залежно від обсягу вибірки відбирають для дослідження кожен 2, 3, 4, 10-ту і т. д. одиницю.

Механічну вибірку широко застосовують для контролю якості продукції, відбору підприємств певної галузі для дослідження, під час бюджетних обстежень сімей.

**Типова (районована) вибірка.** Досліджувану генеральну сукупність розбивають на однорідні групи, райони чи зони. Потім із кожної групи випадково відбирають певну кількість одиниць пропорційно частці цієї групи в загальній сукупності.

Наприклад, відбираючи сім'ї для бюджетного обстеження, їх групують за видами економічної діяльності, а потім — за галузями промисловості, виробничими напрямками тощо. Така процедура забезпечує потрапляння у вибірку представників усіх виділених типових груп, і вибірка стає більш достовірною.

**Серійна (гніздова) вибірка.** На практиці іноді замість відбору окремих одиниць відбирають цілі групи (гнізда), і вже в кожній групі спостерігають усі без винятку одиниці. Групи (гнізда) відбирають методом власне випадкової неповторної вибірки чи за допомогою механічного відбору.

Наприклад, з інкасаторської сумки випадково відбирають 10 пачок грошей по 100 купюр (це гнізда), а потім перевіряють кожен купюру на зношеність або автентичність.

### ***3. Методи формування випадкових вибірок***

Найважливішою властивістю статистичної обробки даних є проблема репрезентативності вибірки. Вибірка повинна адекватно відображати властивості генеральної сукупності. Виникає закономірне питання, як сформуванню репрезентативну вибірку? З точки зору статистики репрезентативність вибірки означає, що представлений в вибірці розподіл досліджуваних ознак відповідає (з певною часткою похибки) їх розподілу в генеральній сукупності. Наведемо два методи, що забезпечують репрезентативність вибірки.

**Перший** метод формування простої випадкової вибірки. У цьому випадку вибірка складається з елементів, відібраних з генеральної сукупності таким чином, щоб кожен елемент цієї сукупності мав би рівні можливості (рівну ймовірність) потрапити до вибірки. Отримана таким чином вибірка називається простою випадковою вибіркою.

**Другий** метод ґрунтується на понятті стратифікованої випадкової вибірки. Для цього необхідно розбити елементи генеральної сукупності на страти (групи) за деякими

характеристиками. Наприклад, при обстеженні певної групи людей генеральну сукупність бажано розбити на групи, наприклад за віком, статтю, соціальним статусом тощо. Якщо проведена подібна розбивка сукупності і випадкова вибірка формується окремо з кожної групи (страти), то отримана в результаті вибірка носить назву стратифікована випадкова вибірка.

Як визначається обсяг вибірки? Підкреслимо, що він залежить насамперед від завдань дослідження. Тим не менш, можна сформулювати найбільш загальні рекомендації:

- Найбільший обсяг вибірки необхідний при розробці діагностичної методики – від 200 до 1000-2500 чоловік.

- Якщо необхідно порівнювати дві вибірки, їх загальна чисельність повинна бути не менше 50 осіб; чисельність порівнюваних вибірок повинна бути приблизно однаковою.

- Якщо вивчається взаємозв'язок між будь-якими властивостями, то обсяг вибірки повинен бути не менше 30-35 осіб.

- Чим більша мінливість досліджуваної властивості, тим більший повинен бути обсяг вибірки. Мінливість можна зменшити, збільшуючи однорідність вибірки, наприклад, за статтю, віком і т.д. При цьому, відповідно, зменшуються можливості генералізації висновків.

У практичній діяльності використовують різноманітні способи відбору об'єктів із генеральної сукупності. Усі способи відбору можна поділити на два види:

1. Вибір, який не потребує розділення генеральної сукупності на частини. До цього виду вибору відносять:

- простий випадковий неповторний відбір;
- простий випадковий повторний відбір.

2. Вибір, при якому генеральна сукупність розділяється на частини (розшарований випадковий відбір). До цього виду вибору відносять:

- типовий відбір;
- механічний відбір;
- серійний відбір.

#### **Умови здійснення простої випадкової вибірки.**

Для здійснення простої випадкової вибірки необхідна наявність основи вибірки, тобто такого представлення генеральної сукупності, при якому її елементи були хоча б перераховані. Наведемо приклади основ вибірки.

#### **Приклад 1.**

а) Генеральна сукупність – всі покупці крамниці. Основою вибірки можуть бути робочі списки покупців, що веде крамниця.

б) Генеральна сукупність – всі жителі міста, що мають телефон. Основою вибірки може бути довідкова телефонна книга.

Як правило, дані для утворення випадкової вибірки представляються у вигляді деякої, заздалегідь складеної таблиці і тому основою вибірки є нумерація елементів цієї таблиці. Основа вибірки повинна повністю відбивати признак генеральної сукупності, що

вивчається. Порушення цієї вимоги може зробити вибірку не репрезентативною. Пояснимо це на прикладі.

### **Приклад 2.**

Вимагається обстежити всі молоді сім'ї невеликого міста на предмет кількості дітей дошкільного віку у сім'ях. З цією метою співробітник відділу, що займається цим питанням, випадковим чином за допомогою телефонного довідника обдзвонює сім'ї з 18.00 до 21.00 кожний день. Чи буде вибірка даних, отримана таким чином, репрезентативною?

Основою вибірки у цьому прикладі є телефонний довідник. Проста випадкова вибірка, що складена за цією основою, не буде репрезентативною по цілому ряду причин: не всі сім'ї міста, що обстежуються, мають телефон, деякі сім'ї у цей період не будуть знаходитися вдома, або члени цих сімей не зможуть підійти до телефону, деякі сім'ї користуються телефонами, номери яких не записані в довідник і так далі.

Наведемо приклади, де може бути використана проста випадкова вибірка.

### **Приклад 3.**

а) Телефонна компанія перевіряє рахунки 10% всіх міжнародних телефонних переговорів з метою визначення їхньої середньої величини.

### **Багаторівнева вибірка.**

Багаторівневий вибір передбачає кілька кроків з попереднім розподілом генеральної сукупності на ієрархічно-групову структуру. На першому кроці випадково вибирають кілька груп найвищого рівня, на другому – у кожній з цих груп вибирають кілька підгруп і так до кінця. На останньому кроці вибирають безпосередньо елементи генеральної сукупності.

Скажімо, якщо потрібно здійснити опитування студентів деякого університету, то сукупність загальної чисельності студентів має природну ієрархію: факультети, відділення, курси, академічні групи, студенти. Спочатку випадково вибираємо кілька факультетів, на кожному з них – денне, заочне або вечірнє відділення, далі для кожного відділення випадково вибираємо курс, потім кілька груп, нарешті, у кожній групі випадково опитуємо кількох студентів.

### **Кластерна вибірка**

Під кластерною вибіркою розуміють багаторівневий вибір з генеральної сукупності, ієрархічну структуру якої сформовано за географічною ознакою. Наприклад, при аналізі генеральної сукупності, яка складається із загальної чисельності громадян країни, можна спочатку випадково вибрати кілька міст і сіл, у кожному з них – кілька районів, у кожному районі – кілька вулиць, на кожній вулиці – кілька будинків, у кожному будинку – мешканців кількох квартир. При цьому на кожному кроці використовують звичайну техніку випадкового вибору.

### **Стратифікаційна вибірка**

Припустимо, за деякою ознакою можна поділити генеральну сукупність на кілька груп одного ієрархічного рівня. Припустимо також, що відоме кількісне співвідношення

обсягів цих груп. Тоді вибірку, яка враховує це співвідношення, називають стратифікаційною. Наприклад, нехай відомо, що у вищому навчальному закладі є 200 професорів, 400 доцентів та 600 асистентів. Якщо потрібно здійснити їх опитування, можна сформувати випадкову вибірку, в якій співвідношення професори : доценти : асистенти становитиме  $200 : 400 : 600 = 2 : 4 : 6 = 1 : 2 : 3$ . Скажімо, можна випадково вибрати 3 професора, 6 доцентів і 9 асистентів. У результаті отримаємо випадкову вибірку з 18 осіб, яка враховує співвідношення чисельності категорій викладачів.

Як правило, стратифікаційні вибірки краще репрезентують генеральну сукупність, ніж звичайні випадкові. На жаль, інформація про обсяги страт часто невідома або недоступна.

### **Зважування**

Якщо відомі пропорції розподілу генеральної сукупності на страти, їх можна врахувати, навіть здійснивши звичайний випадковий вибір. Для цього використовують так звану процедуру зважування. Найлегше її зрозуміти на прикладі.

Нехай відомо, що в генеральній сукупності співвідношення чоловіків до жінок становить 51 : 49. Припустимо, сформовано випадкову вибірку зі 100 індивідумів з метою встановлення рівня середньої заробітної плати в сукупності. Якщо відомо, що чоловіки в середньому отримують більші заробітні плати, ніж жінки, а в розглядуваній вибірці 60 чоловіків і 40 жінок, то природно очікувати, що вибірка встановить завищений рівень середньої заробітної плати. У такому разі можна застосувати процедуру зважування. Для цього розмір заробітної плати кожного чоловіка потрібно помножити на коригуючий множник

$$\frac{40 \cdot 51}{60 \cdot 49},$$

а для кожної жінки — на

$$\frac{60 \cdot 49}{40 \cdot 51},$$

Як правило, зважування вибірки дає гірші результати, ніж безпосередньо стратифікаційний вибір. Зокрема, зважування може значно впливати на величину помилок. Скажімо, у наведеному прикладі для жінки у вибірці з нетипово великою заробітною платою вона збільшиться ще в 1,5 рази. А в результаті помилка аналізу може значно збільшитись.

### **Ймовірнісний вибір**

Вибірку, ймовірність потрапляння до якої кожного елемента генеральної сукупності відома, називають ймовірнісною. У цьому разі теоретично можна визначити значення помилки. Найважливішою формою ймовірнісної вибірки є така, яка забезпечує однакову ймовірність вибору довільного елемента генеральної сукупності.

*Простою випадковою* називають вибірку, ймовірність вибору якої з генеральної сукупності дорівнює ймовірності вибору довільної вибірки такого самого обсягу.



Власне припущення, що використовується проста випадкова вибірка, покладено в основу більшості теоретичних результатів, які забезпечують можливість опису генеральної сукупності шляхом аналізу вибірки.

Зауважимо, що з позиції теорії просту випадкову вибірку можна згенерувати, використовуючи лише вибирання з поверненням. Іншими словами, навмання вибираємо елемент генеральної сукупності, записуємо інформацію про нього, а далі повертаємо цей елемент у генеральну сукупність. Після цього повторюємо процедуру. При цьому один і той самий елемент може потрапити до вибірки кілька разів.

Практично ж вибірки з поверненням, як правило, не застосовують. Проте можна показати, що для великих генеральних сукупностей і відносно невеликих вибірок вибирання з поверненням і без повернення різняться неістотно. Якщо ж генеральна сукупність невелика, то ймовірність отримати вибірку без повторень при формальному вибиранні з поверненням може бути надто малою. Наприклад, якщо на факультеті 300 студентів і формується вибірка з поверненням обсягом 20, то ймовірність отримати вибірку без повторень становить лише 0,523. Отже, вибирання з поверненням і без повернення в такому разі різняться істотно.

### **Вибірка зі зміненими пропорціями**

Часом з певних міркувань досліднику потрібно у вибірці якнайповніше врахувати характеристики певної підгрупи в генеральній сукупності. Тоді він може свідомо збільшити частку цієї підгрупи у вибірці порівняно з відсотком, який мав би припадати на підгрупу з огляду на її реальний обсяг. Таку вибірку називають вибіркою із зміненими пропорціями.

Наприклад, нехай статистичні дослідження засвідчують, що співвідношення хворих на певну хворобу і здорових людей становить 1:100. Теоретично, якщо потрібно дослідити цю хворобу, слід сформуванати вибірку з такою самою пропорцією. Проте на практиці у вибірку включають набагато більший відсоток хворих.

### **Випадкові експерименти**

Розглянемо тестування нового методу лікування певної хвороби. Для того щоб перевірити ефективність цього методу, виконують таку процедуру. Спочатку з генеральної сукупності всіх хворих-добровольців випадково вибирають групу пацієнтів. Далі цю групу випадково поділяють на дві рівновеликі підгрупи. В одній підгрупі – експериментальній – апробовують нову методику, а у другій – контрольній – імітують апробацію нової методики, використовуючи, наприклад, несправжні пігулки. Порівняння перебігу хвороби у двох підгрупах дає змогу оцінити ефективність методики. Такі процедури називають випадковими експериментами.