

# Кібернетика та системний аналіз

УДК 681.3.06

В.П. Авраменко, В.П. Ткаченко, О.В. Попов

Харківський національний університет радіоелектроніки, Харків

## ФОРМУВАННЯ РЕПРЕЗЕНТАТИВНИХ ВИБІРОК ДЛЯ СТАТИСТИЧНИХ ДОСЛІДЖЕНЬ

*В роботі досліджується метод теорії вибірки, котрий вивчає види вибірок з існуючих сукупностей об'єктів, розробляє моделі та обчислювані процедури побудови вибірок, забезпечує репрезентативність вибірок, вивчає вплив вибірки на якість отриманих результатів та інші аспекти інформативної статистики.*

**Ключові слова:** вибірка, інформативна статистика, репрезентативність.

### 1. Вступ

*1.1. Поняття випадкового вибору.* Значна частина статистичних досліджень пов'язана з описом великих сукупностей об'єктів. Якщо цікавлюча нас сукупність дуже чисельна або її елементи малодоступні, або існують інші причини, що не дозволяють вивчати відразу всі її елементи, то звертаються до вивчення якоїсь частини цієї сукупності.

Така частина сукупності, вибрана для повного дослідження усіх елементів, називається вибіркою або вибірковою сукупністю, а уся множина досліджуваних елементів називається генеральною сукупністю. Бажано сформувати вибірку таким чином, щоб вона щонайкраще представляла б усю генеральну сукупність, тобто була б репрезентативною.

Якщо генеральна сукупність мало відома або зовсім невідома, то для формування репрезентативної вибірки не можна запропонувати нічого кращого, ніж чисто випадковий вибір, який можна почати з випадкового вибору одного об'єкту. Вибір одного об'єкту вважається чисто випадковим, якщо всі об'єкти мають рівні імовірності бути вибраними.

При виборі одного об'єкту з  $N$  вірогідність вибору кожного елементу рівна  $1/N$ . Вибір  $n$  об'єктів з  $N$  вважається чисто випадковим, якщо всі набори з  $n$  об'єктів мають однакову вірогідність бути вибраними.

Чисто випадковий вибір  $n$  об'єктів (випадкову вибірку об'єму  $n$ ) можна отримати, витягуючи з генеральної сукупності по одному об'єкту послідовно і чисто випадково. Порушення принципів випадкового вибору може привести до непрезентативності вибірки.

*1.2. Сутність статистичних досліджень.* Статистичне дослідження може бути повним і вибіркоvim. При повному дослідженні вимірювання досліджуваної ознаки проводиться для кожного елементу сукупності. При цьому визначається точний розподіл ознаки. Наприклад, якщо декан отримав розподіл оцінок на екзамені для усіх 200 студентів, то він може визначити долі відмінників і невстигаючих, відсоток

успішності, відсоток якості навчання і тому подібне. Але це ще не «справжня» статистика.

«Справжня» статистика ставить і розв'язує задачу, як, обстеживши елементи вибірки із генеральної сукупності, отримати необхідну інформацію про генеральну сукупність. Перше, що повинен вирішити статистик – це як сформувати вибірку, щоб вона найкращим чином відповідала генеральній сукупності, тобто, щоб вибірка була репрезентативною.

Вибірка буде репрезентативною, якщо відбір елементів у вибірку проводиться випадково. Це означає, що всі елементи генеральної сукупності мають однакову вірогідність потрапити у вибірку. Для отримання випадкової вибірки слід скористатися відповідною програмою ЕОМ.

Забезпечити випадковість вибірки досить не просто. Не можна полагатися на власну інтуїцію. Якщо вибірка виявиться не репрезентативною, то зростанні її обсягу може зменшуватися точність або можуть з'являтися помилкові висновки. Закон великих чисел починає працювати зворотно.

Таким чином, для того щоб провести обстеження великої сукупності (наприклад, усіх студентів України), необхідно зробити вибірку із великої початкової сукупності (наприклад, усіх студентів міста Харкова) і обстежити цю вибірку. Після цього необхідно вирішити питання, як інформація, отримана від вибірки (харків'ян), може бути використана для вивчення великої початкової сукупності (усієї країни).

*1.3. Генеральна сукупність і вибірка.* Генеральною сукупністю (або сукупністю) вважається множина усіх об'єктів, властивості яких цікавлять дослідника. Вибіркою прийнято вважати підмножину генеральної сукупності, яка реально вивчається. Вибірка повинна адекватно відображати властивості сукупності, що цікавлять дослідника. Результати вибіркового дослідження мають бути поширені на усю сукупність [1, с. 151].

У більшості статистичних досліджень потрібно з деякої генеральної сукупності сформувати вибірку згідно заданим показникам адекватності. На підставі

властивостей випадкової вибірки і висунутого критерію ефективності необхідно оцінити вибіркові статистичні характеристики. По вибірковим характеристикам належить методом математичної індукції зробити висновок про характеристики усієї генеральної сукупності досліджуваних об'єктів.

Таким чином, множина усіх можливих уявних спостережень, які могли б бути проведені при даному комплексі умов, прийнято вважати генеральною сукупністю. Оскільки у визначенні йдеться про уявно можливі спостереження, то генеральна сукупність вважається абстрактним поняттям, яке не слід змішувати з реальними сукупностями, що підлягають статистичному дослідженню. Закон або функція розподілу вірогідності повністю характеризують генеральну сукупність.

Вибірку можна тлумачити як деякий емпіричний аналог (модель) генеральної сукупності, з яким найчастіше на практиці доводиться мати справу, оскільки дослідження усієї генеральної сукупності буває або дуже трудомістким, або принципово неможливим. Задачі математичної статистики фактично зводяться до обґрунтування висновку про об'єктивні властивості генеральної сукупності по результатам вибірки.

## 2. Дослідження репрезентативності та непререзентативності вибірок

*2.1. Визначення репрезентативності вибірки.* Достовірність висновків, що отримуються в результаті статистичної обробки даних, у багатьох випадках залежить від успішного вирішення питання репрезентативності вибірки, тобто повноти і адекватності подання властивостей генеральної сукупності моделлю у вигляді вибірки [2, с. 18].

Співвідношення між властивостями генеральної сукупності і характеристиками вибірки можуть задовольняти умовам «репрезентативності», а можуть і не задовольняти їм. Слово «репрезентативний» означає типовий, характерний, адекватний, представлений рисами, подібний до чого-небудь.

Репрезентативність досягається завдяки побудові вибірки, яка найкращим чином представляє генеральну сукупність і дозволяє обґрунтовано переносити висновки, отримані при аналізі вибірки, на генеральну сукупність.

Репрезентативність являє собою таке подання інформації у вигляді, достатньому для отримання обґрунтованого розв'язку, заради якого воно (представлення) було створено. Наприклад, у вибіркового методі статистики репрезентативність інформації являє собою відповідність (адекватність) характеристик вибірки характеристикам генеральної сукупності.

Таким чином, найважливішою властивістю статистичної обробки даних є проблема репрезентативності вибірки. Вибірка повинна адекватно відображати властивості генеральної сукупності. Репрезентативна вибірка гомоморфна генеральній сукупності (гомоморфізм – це відображення, яке зберігає базові

операції та відношення). В забезпеченні репрезентативності вибірок слід виходити з вивчення властивостей генеральної сукупності.

Репрезентативність вибірки досягається випадковістю відбору, коли кожний елемент генеральної сукупності має однакову ймовірність бути відібраним. При оцінюванні репрезентативності вибірки враховується те, яким чином отримана вибірка, і те, яку кількість показників генеральної сукупності «перекривають» показники вибірки.

*2.2. Забезпечення репрезентативності вибірки.* Випадкові похибки вибірових оцінок визначаються відношеннями дисперсії випадкової величини в генеральній сукупності до обсягу вибірки.

Оскільки генеральна дисперсія зазвичай буває невідома, то випадкові похибки приблизно обчислюються через дисперсію вибірки [1, с. 158]:

$$m_p = \sqrt{p(1-p)/n}, \quad m_m = \sqrt{D_x/n}, \quad m_\sigma = \sqrt{D_x/2n}, \\ m_r = \sqrt{(1-r^2)/(n-2)}, \quad m_z = \sqrt{1/(n-3)};$$

тут  $m_i$  при  $i = p, m, \sigma, r, z$  – означає випадкову похибку оцінки вірогідності, арифметичного середнього, стандартного відхилення, коефіцієнта лінійної кореляції і його  $z$ -перетворення; у чисельниках відношень записані відповідні дисперсії,  $n$  – обсяг вибірки.

Відзначені похибки являють собою стандартні відхилення розподілів вибірових випадкових оцінок в ідеальній генеральній сукупності. Для реальних сукупностей ці формули можуть не підходити, оскільки більшість реальних сукупностей мають напіввипадкові вибірки.

Щоб сконструювати таку вибірку, необхідно мати модель оригіналу, тобто модель генеральної сукупності. Тому розробка моделі сукупності повинна передувати проектуванню репрезентативної вибірки або множини статистично однорідних вибірок.

Моделлю складної генеральної сукупності випадкових об'єктів може бути тільки повний скінченномірний розподіл вірогідності  $P(A)$  або  $P(X)$ , де  $A$  або  $X$  повні групи сумісних подій або величин, котрі якісно і кількісно характеризують властивості сукупності, що цікавлять дослідника. Мірність таких розподілів може змінюватися від одиниці до семи окремих властивостей або якостей, що описують сукупність. Але чим вище мірність моделі, тим більшим має бути обсяг вибірки.

*2.3. Причини вибіркового обстеження.* Розв'язання практичних задач проведення вибірових обстежень пов'язано з такими причинами [4, с. 19]:

– при контролі якості продукції великосерійного і масового виробництва генеральна сукупність буває настільки чисельна, що дослідження усіх елементів сукупності (повний перебір об'єктів) стає занадто трудомістким;

– при нескінченно великій генеральній сукупності, коли навіть дуже велика множина спостережень не вичерпує усю генеральну сукупність, оскільки вибірові висновки, отримані по результатах скінченного

числа спостережень, мають бути поширені на усю генеральну сукупність, яка охоплює навіть ту частину об'єктів, які у майбутньому будуть виготовлені на протязі наступних декількох років;

– при проведенні руйнівного контролю, коли в процесі проведення випробувань здійснюється руйнація відібраних зразків (наприклад, випробовування границі міцності тощо);

– іноді можуть існувати результати випробувань усієї сукупності, реально існуючої на даний момент, але за змістом задачі ці результати доцільно розглядати як вибірку із гіпотетичної генеральної сукупності. Так поступають у випадках, коли бажають виявити загальну закономірність, по відношенню до якої існуюча сукупність виступає лише окремим випадком.

**2.4. Класифікація вибірових сукупностей.** В основу класифікації виборок з генеральної сукупності можуть бути покладені такі ознаки: спосіб обробки, обсяг та вимірність вибірки, тривалість використання, єдиність або множинність, національна приналежність [1, с. 154].

За способом відбору слід розрізнити випадкові, не випадкові і напіввипадкові, неповторні і повторні, одноразові і багаторазові, механічні, серійні і типові вибірки. Причому способи відбору можуть комбінуватися.

Випадкова вибірка утворюється випадковим відбором елементів сукупності. Це необхідно для того, щоб уникнути ненавмисного або свідомого спотворення фактів.

Невипадкова, тенденційна, вибірка якраз і утворюється навмисним вибором елементів сукупності, що, звичайно, спотворює образ реальності.

Напіввипадкова вибірка утворюється поєднанням не випадкового і випадкового відбору, що необхідне для репрезентативного відображення складних реальних сукупностей.

Бесповторний відбір, при якому одного разу вибрані елементи не можуть бути вибраними знову і знову, створює фіктивну залежність при малому обсязі генеральної сукупності.

Наприклад, якщо вчитель не питає повторно учнів, що одного разу відповіли на уроці, то вірогідність бути запитаним у решти учнів весь час збільшується. А при повторному опитуванні вірогідність бути запитаним на уроці у всього класу зберігається незмінною.

Повторний відбір означає, що елемент, вибраний одного разу у вибірку, може бути вибраний і в інші такі ж вибірки. Відзначимо, що повторний або бесповторний відбір називають «схемою незалежних або залежних випробувань».

До напіввипадкових вибірок належать механічні, серійні і типові вибірки. При механічному відборі генеральну сукупність розділяють на частини по числу елементів (обсягу) вибірки, а потім з кожної частини по жеребу вибирають один елемент.

При серійному відборі генеральну сукупність поділяють на множину невеликих серій, з яких випадковим відбором формують вибірку серій.

При типовому відборі сукупність поділяється на гомогенні (однорідні по своєму складу) частини, з яких формуються випадкові підвибірки.

Обсяг вибірки, тобто число вибраних для вивчення елементів сукупності, вважається найважливішою властивістю вибірки: від нього прямо або побічно залежить репрезентативність вибірки.

**2.5. Критичний обсяг вибірки.** За об'ємом розрізняють малі, середні і великі вибірки. За малі вважаються вибірки від двох до тридцяти елементів; великі вибірки охоплюють від тисячі до декількох десятків тисяч і більш за елементи.

У вибірку малого обсягу рідкісні елементи сукупності потрапити при випадковому відборі не можуть. З іншого боку, чим більше обсяг вибірки, тим більше витрати потрібні на створення і вивчення вибірки. Тому вибірка за обсягом має бути не тільки необхідною і достатньою, але і економічною.

Необхідний, достатній і економічний обсяг вибірки називається критичним обсягом. У будь-якому вибіровому дослідженні завжди виникає проблема визначення величини критичного обсягу вибірок. Відмітимо, що доцільно мати декілька вибірок невеликого обсягу, котрі «перевіряють одна одну», чим одну вибірку великого розміру.

По тривалості використання розрізняються одноразові і багаторазові вибірки. Одноразові вибірки застосовуються, наприклад, для переписів населення і для стандартизації психологічних текстів, а багаторазові – для вивчення громадської думки. Багаторазові вибірки називаються «панельні-ми». За обсягом вони можуть складати від 1,5 тис. до 25 і більше тис. респондентів і обов'язково мають резерв для поповнення вибуваючих членів.

## Висновок

Відмінності між вибітками і генеральними сукупностями породжують в науці статистичні проблеми – забезпечення і оцінювання репрезентативності досліджень, статистичну перевірку наукових гіпотез, статистичне виявлення і оцінювання зовнішніх і внутрішніх причин і закономірностей, статистично обґрунтоване планування ефективного експерименту і так далі.

При всій різноманітності вибірок саме випадкові вибірки слугують основою для всіх інших вибірок і виступають основним інструментом об'єктивного дослідження більшості генеральних сукупностей.

При цьому будь-яка випадкова вибірка створює випадковий образ генеральної сукупності. З цієї причини, не дивлячись на те, що і генеральна сукупність і вибірка з неї описуються одними і тими ж кількісними характеристиками – розподілами частот і вірогідностями, їх функціями і числовими параметрами, –

ці кількісні характеристики принципово розрізняються в поняттях і позначеннях. Вибіркові характеристики називаються вибірковими оцінками характеристик генеральної сукупності.

Так, вибірковий розподіл послуговує оцінкою генерального розподілу. І якщо остання (оцінка генерального розподілу) розглядається як теоретичне, то перший (вибірковий розподіл) розглядається як емпіричний розподіл.

Числові параметри вибірки інтерпретуються як вибіркові оцінки параметрів сукупності: вибіркове середнє арифметичне вважається за оцінку математичного очікування сукупності, вибіркова дисперсія – оцінкою генеральної дисперсії і так далі. Відмітимо, що обчислювальні формули суть одні і ті ж. Це створює двозначність в поняттях: вибіркові характеристики є параметрами вибірки і одночасно оцінками параметрів сукупності.

На жаль, ця подвійна роль вибіркових кількісних характеристик не завжди правильно осмислюється і використовується. Суть справи в тому, що як параметри вибірки її числові характеристики суть константи, тобто постійні величини. А як оцінки параметрів генеральної сукупності, ці ж характеристики є величинами змінними і при тому випадковими змінними.

З однієї сукупності шляхом зворотного випадкового відбору можна отримати значно більшу кількість випадкових вибірок порівняльно невеликого обсягу у порівнянні з обсягом генеральної сукупності.

### Список літератури

1. Суходольский, Г.В. Математика для гуманитариев [Текст] / Г.В. Суходольский. – Х.: Изд-во Гуманитарный центр, 2007. – 256 с.
2. Дубров, А.Г. Многомерные статистические методы [Текст] / А.Г. Дубров, В. С. Мхитарян, Л. И. Трошин. – М.: Финансы и статистика, 2000. – 352 с.
3. Роганов, Е.А. Математика и информатика для юристов: учебник [Текст] / Е. А. Роганов, Н.Б. Тихомиров, А.М. Шелехов. - М.: МГИУ, 2005. - 364 с.
4. Сигел, Э.Ю. Практическая бизнес-статистика ТЕКСТ [Текст]/ Э.Ю. Сигел. – М.: Издательский дом "Вильямс", 2002. – 1056 с.
5. Тюрин, Ю.Н. Анализ данных на компьютере [Текст] / Ю.Н. Тюрин, А.Н. Макаров. – М.: ИНФРА-М, 2003. – 544 с.
6. Изуменцева, Н.В. Статистический анализ результатов и наблюдений: [Текст] учебн. пос. / Н.В. Изуменцева, В.И. Пахомов. – Х.: «Компания СМІТ», 2005. – 236 с.

Надійшла до редколегії 19.04.2011

**Рецензент:** д-р техн. наук, проф. І.В. Рубан, Харківський університет Повітряних Сил ім. І. Кожедуба, Харків.

### ФОРМИРОВАНИЕ РЕПРЕЗЕНТАТИВНЫХ ВЫБОРОК ДЛЯ СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЙ

В.П. Авраменко, В.П. Ткаченко, А.В. Попов

*В работе исследуется метод теории выборок, который изучает виды выборок из существующей совокупности объектов, разрабатывает модели и вычислительные процедуры построения выборок, обеспечивает репрезентативность выборок, изучает влияние выборки на качество полученных результатов и другие аспекты информативной статистики.*

**Ключевые слова:** выборка, информативная статистика, репрезентативность.

### FORMING OF REPRESENTATIVENESS SAMPLING FOR STATISTICAL RESEARCHES

V.P. Avramenko, V.P. Tkachenko, A.V. Priests

*The method of sampling selections theory is in-process probed, which studies the types of sampling from the existent aggregate of objects, develops models and calculable procedures of construction of sampling, provides representativeness of samplings, studies influence of sampling on quality of the got results and other aspects of informing statistics.*

**Keywords:** selection, informing statistics, representativeness.