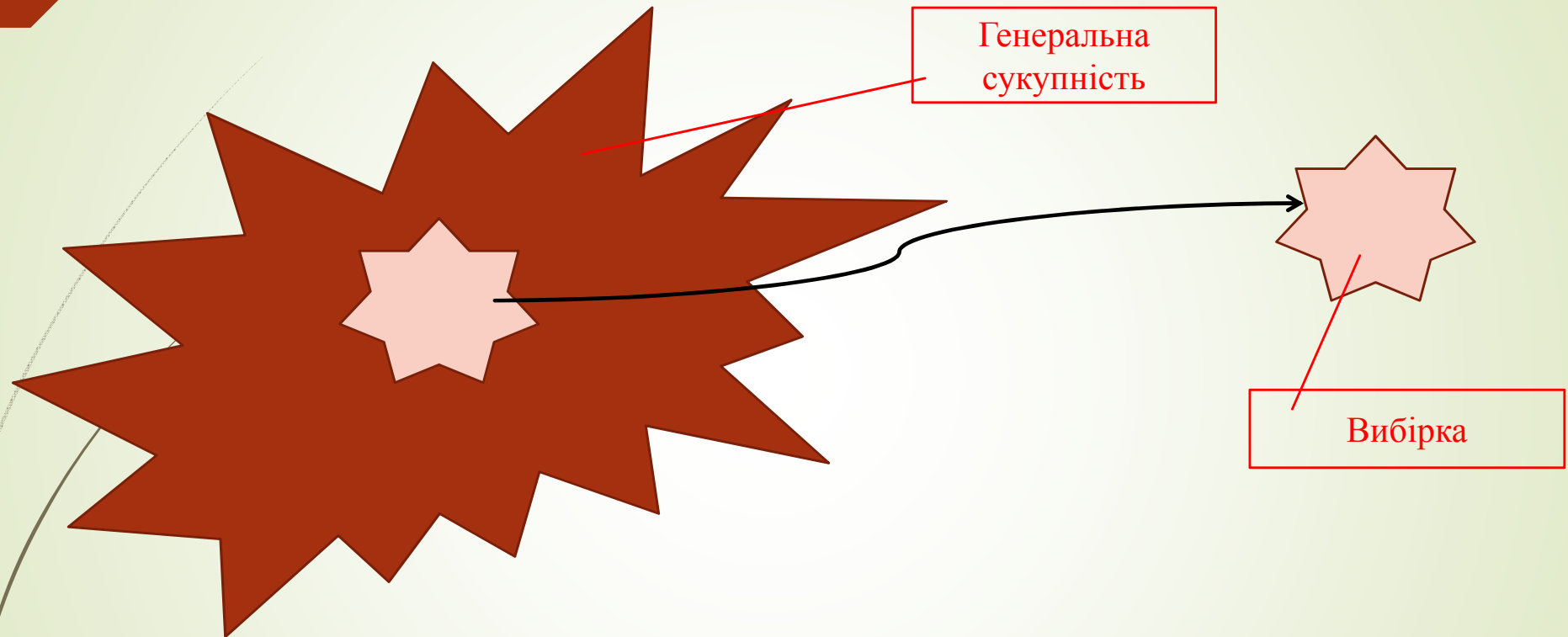


**Структури даних.
Класифікація,
розподіл та
перетворення даних**

Етапи статистичного аналізу:

- ▶ планування збору даних;
- ▶ попереднє дослідження даних;
- ▶ оцінка невідомої величини;
- ▶ перевірка гіпотез.


Планування збору даних :



Процес узагальнення результатів дослідження вибірки на усю сукупність називається **статистичним висновком**.

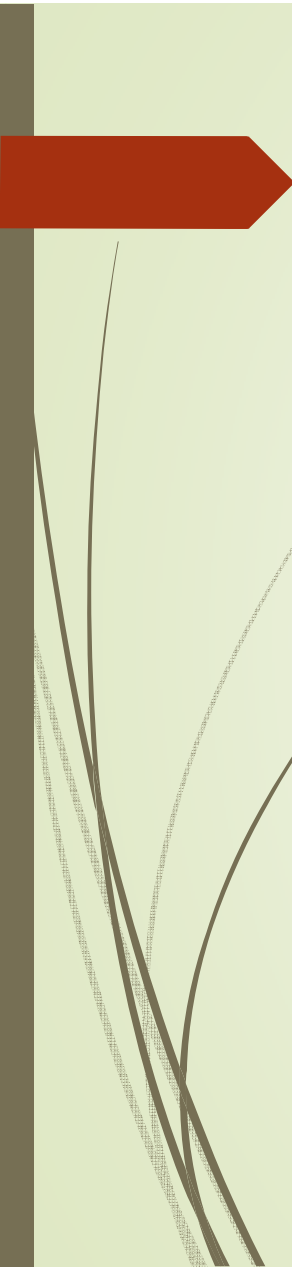
Випадкове вилучення вибірки досягає дві мети:

- гарантування, що процес вилучення вибірки є безпристрасним, тобто всі об'єкти генеральної сукупності мають рівні шанси бути відібраними (вибірка є репрезентативною для даної генеральної сукупності);
- випадковість, введена контрольованим способом на стадії планування проекту, гарантує валідність (коректність) наступних статистичних висновків.



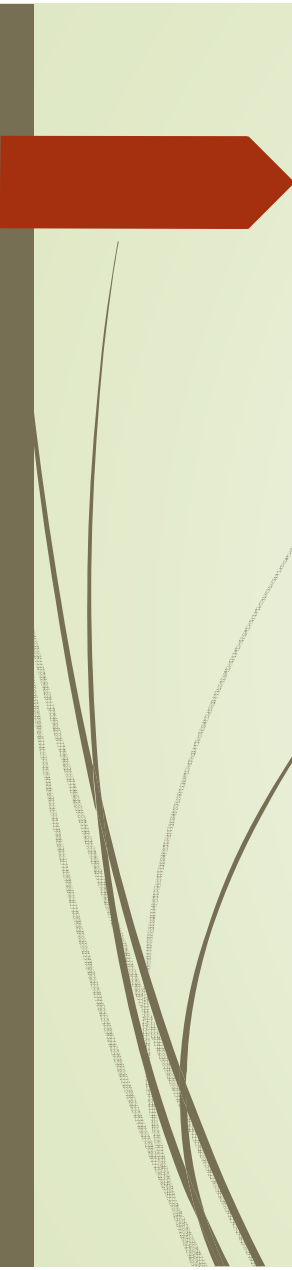
Попереднє дослідження даних – дані аналізуються з різних точок зору, описуються та узагальнюються. Добре виконаний попередній аналіз готує дослідника до проведення формального аналізу:

- ▶ шляхом перевірки, що очікувані зв'язки дійсно існують в даних, й, таким чином, заплановані методи аналізу адекватні даним;
- ▶ шляхом виявлення у даних неочікуваної структури, яку необхідно взяти до уваги, що передбачає внесення змін до плану аналізу.




Оцінка невідомої величини – найбільш обґрунтоване припущення, яке базується на наявних даних. Тому бажано оцінювати ті параметри, які неможливо оцінити точно. Якщо відомо, наскільки (наближено) точні оцінки, то можна вирішити, якою мірою їх слід брати до уваги.

Довірчий інтервал дає ймовірне значення верхньої та нижньої меж невідомої величини, що оцінюється. Обчислення довірчого інтервалу показує наскільки надійною є оцінка насправді.



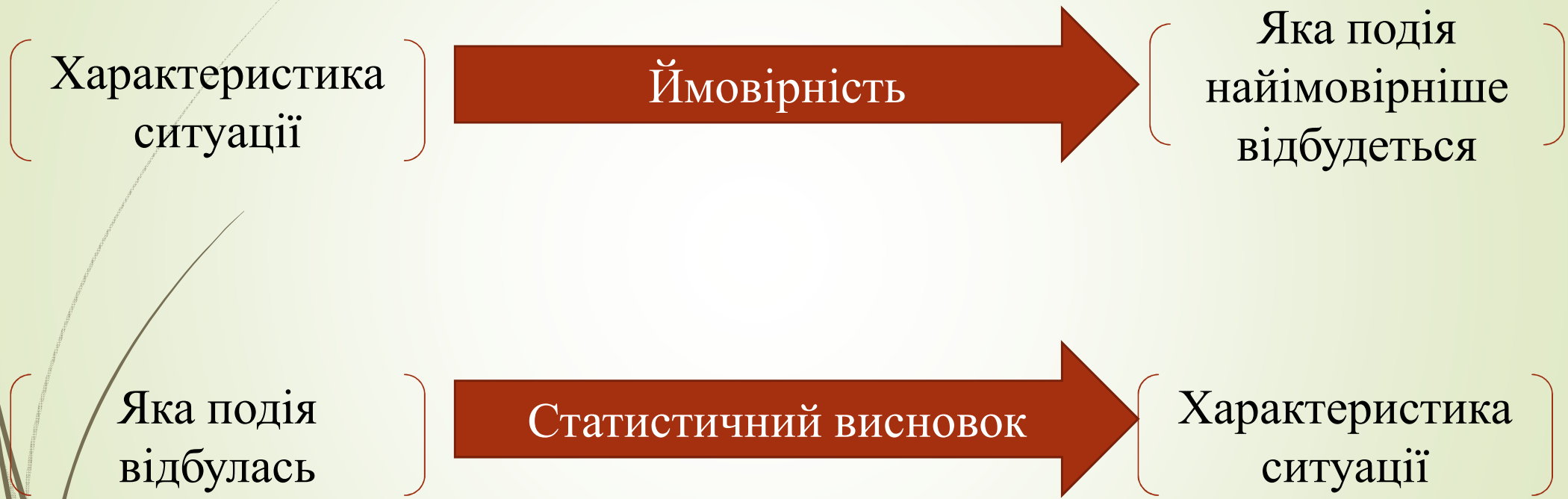
Перевірка статистичних гіпотез полягає у використанні даних для здійснення вибору однієї з двох (або більше) різних можливостей при вирішенні питання у неоднозначній ситуації. Перевірка гіпотези виходячи з даних дає певне рішення про те, яка з можливостей є вірною. *Процедура перевірки гіпотези* включає збір даних, які дозволяють обрати одну з можливостей, й використання статистичного аналізу для підтвердження прийнятого рішення, якщо це рішення не є результатом простого поверхневого аналізу даних.




Ймовірність – це засіб для роботи з ризиком та невизначеністю.

Ймовірність показує можливість (або шанс) настання у майбутньому кожної з різних потенціальних подій, розраховану на основі інформації про певну ситуацію. Обчислюється ймовірність отримання різноманітних результатів для різних стратегій, щоб побачити переваги кожної з них.

Відмінність між ймовірністю та статистичним висновком





Набір даних складається із результатів спостережень об'єктів, які зазвичай включають реєстрацію однієї й тієї ж самої інформації для кожного об'єкта.

Елементарні одиниці - самі об'єкти (компанії, люди, домогосподарства тощо), щоб відрізнити їх від результатів спостережень та вимірювань (обсяги продажів, прибуток населення, вага тощо).

.

Основні способи класифікації наборів даних:

- ▶ за кількістю порції інформації (змінних) для кожної елементарної одиниці;
- ▶ за типом вимірювання (числа або категорії) для кожного спостереження;
- ▶ за необхідністю впорядкованості у часі записів про вимірювання;
- ▶ за тим, чи збиралась інформація про вимірювання спеціально для цього дослідження або дані збиралися раніше кимось іншим для власних потреб.



Змінна – порція інформації, що реєструється для кожного об'єкта.

В залежності від кількості змінних набори даних діляться на:

- **одномірні;**
- **двомірні;**
- **багатомірні.**

Кількісні дані - числа:

- дискретні кількісні дані;
- неперервні кількісні дані.

Дискретна змінна – це така змінна, яка може приймати значення тільки з певного переліку визначених чисел.

Неперервна змінна – будь-яка числова змінна, яка не є дискретною.

Якісні дані - категорії:

- ▶ порядкові (ординальні) дані;
- ▶ номінальні дані.

Ординальний набір даних – це набір даних, для якого існує змістовний порядок (можливість ранжування).

Номінальні якісні дані визначаються у термінах категорій, які неможливо змістовно впорядкувати.

Часовий ряд та один часовий зріз

Часовий ряд – порядок запису значень даних має змістовний зміст.

Один часовий зріз – послідовність, у якій записані дані, не важлива (впорядкованість у часі відсутня). Мова йде про інформацію про деякі об'єкти у певний момент часу.

Джерела даних

- ▶ первинні дані

(самостійний план збору даних);

- ▶ вторинні дані

(зібрані раніше іншими та для інших цілей).