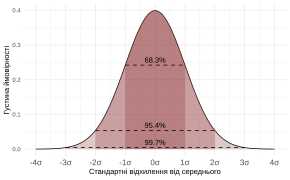
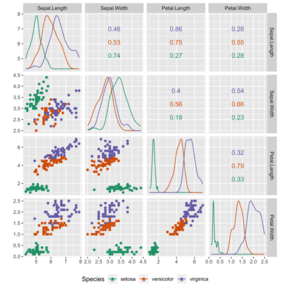
# Статистика

Матеріал з Вікіпедії — вільної енциклопедії.

*Ця стаття про наукову дисципліну. Про чисельну функцію від вибірки див.* [*Статистика (математика)*](https://uk.wikipedia.org/wiki/Статистика_(математика))*.*

*Сюди перенаправляється запит «*[*Статистик*](https://uk.wikipedia.org/w/index.php?title=Статистик&redirect=no)*». На цю тему потрібна* [*окрема стаття*](https://uk.wikipedia.org/wiki/Статистик?action=edit)*.*

|  |
| --- |
| Статистика |
|  |
| * [Нарис](https://uk.wikipedia.org/wiki/Нарис_теорії_статистики) * [Статистики](https://uk.wikipedia.org/w/index.php?title=Список_статистиків&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_statisticians) * [Глосарій](https://uk.wikipedia.org/w/index.php?title=Глосарій_теорії_ймовірностей_та_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Glossary_of_probability_and_statistics) * [Позначення](https://uk.wikipedia.org/w/index.php?title=Позначення_в_теорії_ймовірностей_та_статистиці&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Notation_in_probability_and_statistics) * [Журнали](https://uk.wikipedia.org/w/index.php?title=Список_журналів_зі_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_statistics_journals) * [Списки тем](https://uk.wikipedia.org/w/index.php?title=Списки_статистичних_тем&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Lists_of_statistics_topics) * [Статті](https://uk.wikipedia.org/w/index.php?title=Список_статей_зі_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_statistics_articles) * [Статистика](https://uk.wikipedia.org/wiki/Категорія:Статистика) * [Портал «Математика»](https://uk.wikipedia.org/wiki/Портал:Математика) |
|  |

[](https://uk.wikipedia.org/wiki/Файл:Standard_Normal_Distribution_uk.svg)[Нормальний розподіл](https://uk.wikipedia.org/wiki/Нормальний_розподіл), дуже поширена [густина ймовірності](https://uk.wikipedia.org/wiki/Густина_ймовірності), корисна через [центральну граничну теорему](https://uk.wikipedia.org/wiki/Центральна_гранична_теорема). [](https://uk.wikipedia.org/wiki/Файл:Iris_Pairs_Plot.png)[Точкові діаграми](https://uk.wikipedia.org/wiki/Точкова_діаграма) в описовій статистиці використовують, щоби показувати спостережувані взаємозв'язки між різними змінними, тут із застосуванням [набору даних ірисів Фішера](https://uk.wikipedia.org/wiki/Іриси_Фішера).

**Стати́стика** ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistics*) — це дисципліна, що працює над збиранням, організуванням, аналізом, інтерпретуванням та представленням [даних](https://uk.wikipedia.org/wiki/Дані).[[1]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-ox-1)[[2]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-2)[[3]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-3)[[4]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-ЮЕ-4) В застосуванні статистики до наукової, промислової або соціальної задачі є звичним починати з [генеральної сукупності](https://uk.wikipedia.org/wiki/Генеральна_сукупність) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistical population*) або [статистичної моделі](https://uk.wikipedia.org/wiki/Статистична_модель) для дослідження. Генеральні сукупності можуть бути різноманітними групами людей або об'єктів, такими як «всі люди, що живуть в якійсь країні» або «кожен з атомів, що складають кристал». Статистика працює зі всіма аспектами даних, включно з плануванням збирання даних в термінах планування [обстежень](https://uk.wikipedia.org/wiki/Статистичне_обстежування) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *surveys*) та [експериментів](https://uk.wikipedia.org/wiki/Планування_експерименту).[[5]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Dodge-5) Див. [глосарій теорії ймовірностей та статистики](https://uk.wikipedia.org/w/index.php?title=Глосарій_теорії_ймовірностей_та_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Glossary_of_probability_and_statistics).

Коли зібрати дані [перепису](https://uk.wikipedia.org/wiki/Перепис_(статистика)) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *census*) неможливо, [статистики](https://uk.wikipedia.org/wiki/Статистик) збирають дані, розробляючи спеціальні плани експериментів ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *experiment designs*), та [вибірки](https://uk.wikipedia.org/wiki/Вибірка) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *samples*) для обстежування. Репрезентативне вибирання забезпечує можливість розумного розширення висновків та рішень з цієї вибірки на сукупність в цілому. [Експериментальне дослідження](https://uk.wikipedia.org/wiki/Експериментальне_дослідження) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *experimental study*) включає здійснення вимірювань досліджуваної системи, маніпулювання цією системою, а потім здійснення додаткових вимірювань із застосуванням тієї ж процедури, щоби визначити, чи змінило маніпулювання значення цих вимірювань. На противагу цьому, [спостережне дослідження](https://uk.wikipedia.org/w/index.php?title=Спостережне_дослідження&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Observational_study) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *observational study*) не включає експериментального маніпулювання.

В [аналізі даних](https://uk.wikipedia.org/wiki/Аналіз_даних) використовують два основні статистичні методи: [описову статистику](https://uk.wikipedia.org/wiki/Описова_статистика) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *descriptive statistics*), яка узагальнює дані з вибірки із застосуванням [статистичних індексів](https://uk.wikipedia.org/w/index.php?title=Статистичний_індекс&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Index_(statistics)), таких як [середнє значення](https://uk.wikipedia.org/wiki/Середнє_значення) та [стандартне відхилення](https://uk.wikipedia.org/wiki/Стандартне_відхилення), та [індуктивну статистику](https://uk.wikipedia.org/wiki/Статистичне_висновування) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *inferential statistics*), яка робить висновки з даних, що піддаються випадковій мінливості (наприклад, похибкам спостережень, варіюванню вибірки).[[6]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-LundResearchLtd-6)[[7]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-7) Описова статистика часто найбільше цікавиться двома наборами властивостей *розподілу* (вибірки або загальної сукупності): [*центральна тенденція*](https://uk.wikipedia.org/wiki/Центральна_тенденція) (або *положення*) прагне схарактеризувати центральне або типове значення цього розподілу, тоді як [*дисперсія*](https://uk.wikipedia.org/wiki/Статистична_дисперсія) (або *мінливість*) характеризує міру, до якої члени цього розподілу відхиляються від його центру, та один від одного. Висновування в [математичній статистиці](https://uk.wikipedia.org/wiki/Математична_статистика) здійснюють в рамках [теорії ймовірностей](https://uk.wikipedia.org/wiki/Теорія_ймовірностей), що займається аналізом випадкових явищ.

Стандартна статистична процедура включає збирання даних, що веде до [перевірки взаємозв'язку](https://uk.wikipedia.org/wiki/Перевірка_статистичних_гіпотез) між двома наборами статистичних даних, або набором даних та синтетичними даними, отриманими з ідеалізованої моделі. Для статистичного взаємозв'язку між двома наборами даних пропонують гіпотезу, і порівнюють її, як [альтернативу](https://uk.wikipedia.org/w/index.php?title=Альтернативна_гіпотеза&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Alternative_hypothesis), з ідеалізованою [нульовою гіпотезою](https://uk.wikipedia.org/wiki/Нульова_гіпотеза) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *null hypothesis*) про відсутність взаємозв'язку між цими двома наборами даних. Відхиляння або спростування нульової гіпотези здійснюють із застосуванням статистичних критеріїв ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistical tests*), що кількісно виражають сенс, в якому хибність нульової гіпотези можливо вважати доведеною за наданих даних, які використовують в цій перевірці. Працюючи від нульової гіпотези, розпізнаю́ть два основні види помилок: [помилки I роду](https://uk.wikipedia.org/wiki/Помилки_I_роду) (нульову гіпотезу хибно відхиляють, що дає «хибно позитивне», [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *false positive*), та [помилки II роду](https://uk.wikipedia.org/wiki/Помилки_II_роду) (нульову гіпотезу не вдається відхилити, й справжній взаємозв'язок між сукупностями втрачається, що дає «хибно негативне», [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *false negative*).[[8]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-8) З цією системою виявилося пов'язано численні проблеми: від отримування достатнього розміру вибірки, й до вказування адекватної нульової гіпотези.[[*джерело?*](https://uk.wikipedia.org/wiki/Вікіпедія:Посилання_на_джерела)]

Процеси вимірювання, що породжують статистичні дані, також піддаються помилкам. Багато з цих помилок класифікують як випадкові (шум, [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *noise*) або систематичні ([упередження](https://uk.wikipedia.org/wiki/Упередження_(статистика)), [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *bias*), але можуть траплятися й інші типи помилок (наприклад, промахи, [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *blunder*, такі як коли аналітик повідомляє неправильні одиниці вимірювання). Існування [пропущених даних](https://uk.wikipedia.org/w/index.php?title=Пропущені_дані&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Missing_data) або [цензурування](https://uk.wikipedia.org/w/index.php?title=Цензурування_(статистика)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Censoring_(statistics)) може призводити до упереджених оцінок, й для подолання цих проблем було розроблено спеціальні методики.

Найраніші праці з [імовірності](https://uk.wikipedia.org/wiki/Імовірність) та статистики, статистичних методів, що спираються на [теорію ймовірностей](https://uk.wikipedia.org/wiki/Теорія_ймовірностей), сходять до [арабських математиків](https://uk.wikipedia.org/wiki/Математика_ісламського_середньовіччя) та [криптографів](https://uk.wikipedia.org/w/index.php?title=Криптографи&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Cryptographers), зокрема, [Аль-Халіля](https://uk.wikipedia.org/wiki/Аль-Халіль_аль-Фарагіді) (717—786)[[9]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-LB-9) та [Аль-Кінді](https://uk.wikipedia.org/wiki/Аль-Кінді) (801—873).[[10]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-sim2000-10)[[11]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-ibr1992-11) У XVIII сторіччі статистика також почала сильно живитися з [диференціального та інтегрального числення](https://uk.wikipedia.org/wiki/Диференціальне_та_інтегральне_числення). Останніми роками для вироблення цих критеріїв, таких як описовий аналіз, статистика покладалася більше на статистичне програмне забезпечення.[[12]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-12)

## Зміст

* [1 Вступ](https://uk.wikipedia.org/wiki/Статистика" \l "Вступ)
  + [1.1 Математична статистика](https://uk.wikipedia.org/wiki/Статистика" \l "Математична_статистика)
* [2 Історія](https://uk.wikipedia.org/wiki/Статистика" \l "Історія)
* [3 Статистичні дані](https://uk.wikipedia.org/wiki/Статистика" \l "Статистичні_дані)
  + [3.1 Збирання даних](https://uk.wikipedia.org/wiki/Статистика" \l "Збирання_даних)
    - [3.1.1 Вибирання](https://uk.wikipedia.org/wiki/Статистика" \l "Вибирання)
    - [3.1.2 Експериментальні та спостережні дослідження](https://uk.wikipedia.org/wiki/Статистика" \l "Експериментальні_та_спостережні_дослідження)
      * [3.1.2.1 Експерименти](https://uk.wikipedia.org/wiki/Статистика" \l "Експерименти)
      * [3.1.2.2 Спостережне дослідження](https://uk.wikipedia.org/wiki/Статистика" \l "Спостережне_дослідження)
  + [3.2 Типи даних](https://uk.wikipedia.org/wiki/Статистика" \l "Типи_даних)
* [4 Статистичні методи](https://uk.wikipedia.org/wiki/Статистика" \l "Статистичні_методи)
  + [4.1 Описова статистика](https://uk.wikipedia.org/wiki/Статистика" \l "Описова_статистика)
  + [4.2 Індуктивна статистика](https://uk.wikipedia.org/wiki/Статистика" \l "Індуктивна_статистика)
    - [4.2.1 Термінологія та теорія індуктивної статистики](https://uk.wikipedia.org/wiki/Статистика" \l "Термінологія_та_теорія_індуктивної_статистики)
      * [4.2.1.1 Статистики, статистичні оцінки, та центральні величини](https://uk.wikipedia.org/wiki/Статистика" \l "Статистики,_статистичні_оцінки,_та_центральні_величини)
      * [4.2.1.2 Нульова гіпотеза та альтернативна гіпотеза](https://uk.wikipedia.org/wiki/Статистика" \l "Нульова_гіпотеза_та_альтернативна_гіпотеза)
      * [4.2.1.3 Похибка](https://uk.wikipedia.org/wiki/Статистика" \l "Похибка)
      * [4.2.1.4 Інтервальне оцінювання](https://uk.wikipedia.org/wiki/Статистика" \l "Інтервальне_оцінювання)
      * [4.2.1.5 Значущість](https://uk.wikipedia.org/wiki/Статистика" \l "Значущість)
      * [4.2.1.6 Приклади](https://uk.wikipedia.org/wiki/Статистика" \l "Приклади)
  + [4.3 Розвідувальний аналіз даних](https://uk.wikipedia.org/wiki/Статистика" \l "Розвідувальний_аналіз_даних)
* [5 Неналежне застосування](https://uk.wikipedia.org/wiki/Статистика" \l "Неналежне_застосування)
  + [5.1 Неналежна інтерпретація: кореляція](https://uk.wikipedia.org/wiki/Статистика" \l "Неналежна_інтерпретація:_кореляція)
* [6 Застосування](https://uk.wikipedia.org/wiki/Статистика" \l "Застосування)
  + [6.1 Прикладна статистика, теоретична статистика та математична статистика](https://uk.wikipedia.org/wiki/Статистика" \l "Прикладна_статистика,_теоретична_статистика_та_математична_статистика)
  + [6.2 Машинне навчання та добування даних](https://uk.wikipedia.org/wiki/Статистика" \l "Машинне_навчання_та_добування_даних)
  + [6.3 Статистика в академічнім середовищі](https://uk.wikipedia.org/wiki/Статистика" \l "Статистика_в_академічнім_середовищі)
  + [6.4 Статистичні обчислення](https://uk.wikipedia.org/wiki/Статистика" \l "Статистичні_обчислення)
  + [6.5 Застосування статистики в математиці та мистецтві](https://uk.wikipedia.org/wiki/Статистика" \l "Застосування_статистики_в_математиці_та_мистецтві)
* [7 Спеціалізовані дисципліни](https://uk.wikipedia.org/wiki/Статистика" \l "Спеціалізовані_дисципліни)
* [8 Див. також](https://uk.wikipedia.org/wiki/Статистика" \l "Див._також)
* [9 Література](https://uk.wikipedia.org/wiki/Статистика" \l "Література)
* [10 Посилання](https://uk.wikipedia.org/wiki/Статистика" \l "Посилання)
* [11 Примітки](https://uk.wikipedia.org/wiki/Статистика" \l "Примітки)

## Вступ

*Докладніше:* [*Нарис теорії статистики*](https://uk.wikipedia.org/wiki/Нарис_теорії_статистики)

|  |
| --- |
| Частина серії статей з статистики |
| [Теорія ймовірностей](https://uk.wikipedia.org/wiki/Теорія_ймовірностей) |
|  |
| * [Аксіоми](https://uk.wikipedia.org/wiki/Аксіоматика_теорії_ймовірностей) |
| * [Ймовірнісний простір](https://uk.wikipedia.org/wiki/Ймовірнісний_простір) * [Простір елементарних подій](https://uk.wikipedia.org/wiki/Простір_елементарних_подій) * [Елементарна подія](https://uk.wikipedia.org/wiki/Елементарна_подія) * [Випадкова подія](https://uk.wikipedia.org/wiki/Випадкова_подія) * [Випадкова величина](https://uk.wikipedia.org/wiki/Випадкова_величина) * [Міра ймовірності](https://uk.wikipedia.org/wiki/Міра_ймовірності) |
| * [Доповнювальна подія](https://uk.wikipedia.org/w/index.php?title=Доповнювальна_подія&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Complementary_event) * [Спільний розподіл](https://uk.wikipedia.org/wiki/Спільний_розподіл) * [Відособлений розподіл](https://uk.wikipedia.org/wiki/Відособлений_розподіл) * [Умовна ймовірність](https://uk.wikipedia.org/wiki/Умовна_ймовірність) |
| * [Незалежність](https://uk.wikipedia.org/wiki/Незалежність_(теорія_ймовірностей)) * [Умовна незалежність](https://uk.wikipedia.org/w/index.php?title=Умовна_незалежність&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Conditional_independence) * [Формула повної ймовірності](https://uk.wikipedia.org/wiki/Формула_повної_ймовірності) * [Закон великих чисел](https://uk.wikipedia.org/wiki/Закон_великих_чисел) * [Теорема Баєса](https://uk.wikipedia.org/wiki/Теорема_Баєса) * [Нерівність Буля](https://uk.wikipedia.org/w/index.php?title=Нерівність_Буля&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Boole's_inequality) |
| * [Діаграма Венна](https://uk.wikipedia.org/wiki/Діаграма_Венна) * [Деревна діаграма](https://uk.wikipedia.org/w/index.php?title=Деревна_діаграма_(теорія_ймовірностей)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Tree_diagram_(probability_theory)) |
|  |

Статистика — це математичний масив наукових знань, що стосуються збирання, аналізу, інтерпретування або пояснювання, та представлення [даних](https://uk.wikipedia.org/wiki/Дані),[[13]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-13)[[4]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-ЮЕ-4) або одна з галузей [математики](https://uk.wikipedia.org/wiki/Математика).[[14]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-14) Дехто вважає статистику радше окремою математичною наукою, ніж галуззю математики. В той час як даними користуються багато наукових досліджень, статистика займається використанням даних в контексті невизначеності, та ухвалюванням рішень в умовах невизначеності.[[15]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-15)[[16]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-16)

В застосуванні статистики до задачі звичною практикою є починати з [сукупності](https://uk.wikipedia.org/wiki/Генеральна_сукупність) або процесу для дослідження. Сукупності можуть бути різноманітної тематики, такі як «всі люди, що мешкають у країні» або «кожен з атомів, що складають кристал». В ідеалі статистики збирають дані про всю сукупність (операція, звана [переписом](https://uk.wikipedia.org/wiki/Перепис_(статистика))). Це може бути зорганізовано державними статистичними установами. [*Описову статистику*](https://uk.wikipedia.org/wiki/Описова_статистика) можливо використовувати для узагальнювання даних про сукупність. До чисельних описувачів для [неперервних](https://uk.wikipedia.org/wiki/Неперервний_розподіл_ймовірностей) типів даних (таких, як дохід) належать [середнє значення](https://uk.wikipedia.org/wiki/Середнє_значення) та [стандартне відхилення](https://uk.wikipedia.org/wiki/Стандартне_відхилення), тоді як в термінах описування [категорійних даних](https://uk.wikipedia.org/wiki/Категорійні_дані) (таких, як освіта) кориснішими є частота та відсоток.

Коли перепис є нездійсненним, досліджують обрану підмножину сукупності, звану [вибіркою](https://uk.wikipedia.org/wiki/Вибирання_(статистика)) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *sample*). Щойно визначено вибірку, яка для цієї сукупності є репрезентативною, дані для членів цієї вибірки збирають у спостережній або [експериментальній](https://uk.wikipedia.org/wiki/Експеримент) постановці. Знов-таки, для узагальнювання цих вибіркових даних можливо застосовувати описову статистику. Проте, вибирання вибірки мало елементи випадковості, тож встановлені чисельні описувачі з цієї вибірки залежать від випадковості. Щоби все ж таки робити змістовні висновки про сукупність в цілому, потрібна [*індуктивна статистика*](https://uk.wikipedia.org/wiki/Індуктивна_статистика). Вона використовує закономірності в даних вибірки, щоби робити висновки про представлену нею сукупність, з урахуванням випадковості. Ці висновки можуть набувати вигляду: відповідання на питання «так/ні» ([перевіряння гіпотез](https://uk.wikipedia.org/wiki/Перевіряння_гіпотез), [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *hypothesis testing*), оцінювання чисельних характеристик даних ([оцінювання](https://uk.wikipedia.org/wiki/Теорія_оцінювання), [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *estimation*), описування [пов'язаностей](https://uk.wikipedia.org/wiki/Кореляція) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *associations*) в даних ([кореляція](https://uk.wikipedia.org/wiki/Кореляція)), та моделювання взаємозв'язків всередині даних (наприклад, із застосуванням [регресійного аналізу](https://uk.wikipedia.org/wiki/Регресійний_аналіз)). Висновування може розширюватися до [прогнозування](https://uk.wikipedia.org/wiki/Прогнозування), [передбачування](https://uk.wikipedia.org/wiki/Передбачування), та оцінювання неспостережуваних змінних чи то всередині досліджуваної сукупності, чи пов'язаних із нею; до нього можуть належати [екстраполювання](https://uk.wikipedia.org/wiki/Екстраполювання) та [інтерполювання](https://uk.wikipedia.org/wiki/Інтерполювання) [часових рядів](https://uk.wikipedia.org/wiki/Часовий_ряд) та [просторових даних](https://uk.wikipedia.org/wiki/Просторові_дані), а також [добування даних](https://uk.wikipedia.org/wiki/Добування_даних).

### Математична статистика

*Докладніше:* [*Математична статистика*](https://uk.wikipedia.org/wiki/Математична_статистика)

Математична статистика — це застосування [математики](https://uk.wikipedia.org/wiki/Математика) до статистики. До математичних методик, які для цього використовують, належать [математичний аналіз](https://uk.wikipedia.org/wiki/Математичний_аналіз), [лінійна алгебра](https://uk.wikipedia.org/wiki/Лінійна_алгебра), [стохастичний аналіз](https://uk.wikipedia.org/wiki/Стохастичний_аналіз), [диференціальні рівняння](https://uk.wikipedia.org/wiki/Диференціальні_рівняння) та [теорія ймовірностей теорії міри](https://uk.wikipedia.org/w/index.php?title=Теорія_ймовірностей_теорії_міри&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Measure-theoretic_probability_theory).[[17]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-17)[[18]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-18)

## Історія

[](https://uk.wikipedia.org/wiki/Файл:Jerôme_Cardan.jpg)[Джироламо Кардано](https://uk.wikipedia.org/wiki/Джироламо_Кардано), піонер математики ймовірності.

*Детальніші відомості з цієї теми ви можете знайти в статті* [***Історія статистики***](https://uk.wikipedia.org/w/index.php?title=Історія_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/History_of_statistics) ***та*** [***Засновники статистики***](https://uk.wikipedia.org/w/index.php?title=Засновники_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Founders_of_statistics)*.*

Найраніші праці з [імовірності](https://uk.wikipedia.org/wiki/Імовірність) та статистики сходять до [арабських математиків](https://uk.wikipedia.org/wiki/Математика_ісламського_середньовіччя) та [криптографів](https://uk.wikipedia.org/w/index.php?title=Криптографи&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Cryptographers) [Золотої доби ісламу](https://uk.wikipedia.org/wiki/Золота_доба_ісламу) між VIII та XIII сторіччями. [Аль-Халіль](https://uk.wikipedia.org/wiki/Аль-Халіль_аль-Фарагіді) (717—786) написав [«Книгу таємної мови»](https://uk.wikipedia.org/wiki/Кітаб_аль-Муамма_(книга)), що містить перше застосування [перестановок та комбінацій](https://uk.wikipedia.org/w/index.php?title=Перестановки_та_комбінації&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Permutations_and_combinations) з метою перелічування всіх [арабських](https://uk.wikipedia.org/wiki/Арабська_мова) слів з голосними та без них.[[9]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-LB-9) Найранішою книгою зі статистики є [«Трактат про дешифрування криптографічних повідомлень»](https://uk.wikipedia.org/wiki/Трактат_про_дешифрування_криптографічних_повідомлень), написаний арабським вченим [Аль-Кінді](https://uk.wikipedia.org/wiki/Аль-Кінді) (801—873). У своїй книзі Аль-Кінді навів докладний опис того, як застосовувати статистику та [частотний аналіз](https://uk.wikipedia.org/wiki/Частотний_аналіз_(криптологія)) для розшифровування [зашифрованих](https://uk.wikipedia.org/wiki/Шифрування) повідомлень. Цей текст заклав основи статистики та [криптоаналізу](https://uk.wikipedia.org/wiki/Криптоаналіз).[[10]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-sim2000-10)[[11]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-ibr1992-11) Аль-Кінді також здійснив найраніше з відомих застосування [статистичного висновування](https://uk.wikipedia.org/wiki/Статистичне_висновування), тоді як пізніші арабські криптографи розробили ранні статистичні методи для [розкодовування](https://uk.wikipedia.org/wiki/Код) зашифрованих повідомлень. [Ібн Адлан](https://uk.wikipedia.org/w/index.php?title=Ібн_Адлан&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Ibn_Adlan) (1187—1268) пізніше зробив важливий внесок стосовно застосування в частотному аналізі [розміру вибірки](https://uk.wikipedia.org/wiki/Розмір_вибірки).[[9]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-LB-9)

Найраніші європейські праці зі статистики сходять до 1663 року, публікації [«Природні та політичні спостереження на списках померлих»](https://uk.wikipedia.org/wiki/Природні_та_політичні_спостереження_на_списках_померлих) [Джона Ґраунта](https://uk.wikipedia.org/wiki/Джон_Ґраунт).[[19]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-19) Ранні застосування статистичного мислення оберталися довкола потреб держав ([лат.](https://uk.wikipedia.org/wiki/Латинська_мова) *statum*) ґрунтувати політику на демографічних та економічних даних, звідси [етимологія «стат-»](https://uk.wikipedia.org/w/index.php?title=Історія_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/History_of_statistics" \l "Etymology). На початку XIX сторіччя сфера дисципліни статистики розширилася, включивши збирання та аналіз даних в цілому. Натепер статистику широко застосовують в урядуванні, бізнесі, а також природничих та соціальних науках.

Математичні основи сучасної статистики було закладено в XVII сторіччі з розробкою [Джироламо Карданом](https://uk.wikipedia.org/wiki/Джироламо_Кардано), [Блезом Паскалем](https://uk.wikipedia.org/wiki/Блез_Паскаль) та [П'єром Ферма](https://uk.wikipedia.org/wiki/П'єр_Ферма) [теорії ймовірностей](https://uk.wikipedia.org/wiki/Теорія_ймовірностей). Математична теорія ймовірностей постала з досліджень [ігор випадку](https://uk.wikipedia.org/wiki/Гра_випадку), хоч поняття ймовірності й було вже досліджено в [середньовічному праві](https://uk.wikipedia.org/wiki/Середньовічне_римське_право), та філософами, такими як [Хуан Карамуель](https://uk.wikipedia.org/wiki/Хуан_Карамуель).[[20]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-20) [Метод найменших квадратів](https://uk.wikipedia.org/wiki/Метод_найменших_квадратів) було вперше описано [Адрієном-Марі Лежандром](https://uk.wikipedia.org/wiki/Адрієн-Марі_Лежандр) 1805 року.

[](https://uk.wikipedia.org/wiki/Файл:Karl_Pearson,_1910.jpg)[Карл Пірсон](https://uk.wikipedia.org/wiki/Карл_Пірсон), засновник математичної статистики.

Сучасна галузь статистики виникла наприкінці XIX — початку XX сторіччя в три етапи.[[21]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-21) Першу хвилю, на рубежі сторіч, очолила праця [Френсіса Ґолтена](https://uk.wikipedia.org/wiki/Френсіс_Ґолтен) та [Карла Пірсона](https://uk.wikipedia.org/wiki/Карл_Пірсон), які перетворили статистику на строгу математичну дисципліну, яку використовували для аналізу, і не лише в науці, але також і в промисловості та політиці. До внеску Ґолтена належать введення понять [стандартного відхилення](https://uk.wikipedia.org/wiki/Стандартне_відхилення), [кореляції](https://uk.wikipedia.org/wiki/Кореляція), [регресійного аналізу](https://uk.wikipedia.org/wiki/Регресійний_аналіз), та застосування цих методів до дослідження розмаїття людських характеристик — зросту, ваги, довжини вій тощо.[[22]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Galton1877-22) Пірсон, серед іншого, розробив [коефіцієнт кореляції моменту добутку Пірсона](https://uk.wikipedia.org/wiki/Коефіцієнт_кореляції_моменту_добутку_Пірсона), визначений як момент добутку,[[23]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-23) [метод моментів](https://uk.wikipedia.org/wiki/Метод_моментів) для допасовування розподілів до вибірок, та [розподіл Пірсона](https://uk.wikipedia.org/w/index.php?title=Розподіл_Пірсона&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Pearson_distribution).[[24]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Pearson,_On_the_criterion-24) Ґолтен та Пірсон заснували «[*Biometrika*](https://uk.wikipedia.org/w/index.php?title=Biometrika&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Biometrika)» як перший журнал з математичної статистики та [біостатистики](https://uk.wikipedia.org/wiki/Біостатистика) (яку тоді називали біометрією), й пізніше заснували перший в світі університетський статистичний факультет в [Університетськім коледжі Лондона](https://uk.wikipedia.org/wiki/Університетський_коледж_Лондона).[[25]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-25)

[Рональд Фішер](https://uk.wikipedia.org/wiki/Рональд_Фішер) під час експерименту [«пані дегустує чай»](https://uk.wikipedia.org/wiki/Леді_дегустує_чай) закарбував термін [нульова гіпотеза](https://uk.wikipedia.org/wiki/Нульова_гіпотеза), яку «ніколи не доводять та не встановлюють, але, можливо, спростовують в ході експерименту».[[26]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-26)[[27]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-oed-27)

Початок другій хвилі 1910-х та 20-х років поклав [Вільям Сілі Ґоссет](https://uk.wikipedia.org/w/index.php?title=Вільям_Сілі_Ґоссет&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/William_Sealy_Gosset), й вона досягла своєї кульмінації в осяяннях [Рональда Фішера](https://uk.wikipedia.org/wiki/Рональд_Фішер), який написав підручники, що мали визначити цю академічну дисципліну в університетах по всьому світі. Найважливішими публікаціями Фішера були його засаднича праця 1918 року [«Кореляція між родичами в припущенні менделевої спадковості](https://uk.wikipedia.org/w/index.php?title=Кореляція_між_родичами_в_припущенні_менделевої_спадковості&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/The_Correlation_between_Relatives_on_the_Supposition_of_Mendelian_Inheritance)» (яка була першою, де було застосовано статистичний термін [«дисперсія»](https://uk.wikipedia.org/wiki/Дисперсія_випадкової_величини), [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *variance*), його класична праця 1925 року [«Статистичні методи для дослідників](https://uk.wikipedia.org/w/index.php?title=Статистичні_методи_для_дослідників&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Statistical_Methods_for_Research_Workers)», та його [«Планування експериментів](https://uk.wikipedia.org/w/index.php?title=Планування_експериментів_(книга)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/The_Design_of_Experiments)» 1935 року,[[28]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-28)[[29]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-29)[[30]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-30) де він розробив строгі моделі [планування експериментів](https://uk.wikipedia.org/wiki/Планування_експериментів). Він започаткував поняття [достатності](https://uk.wikipedia.org/wiki/Достатність_(статистика)), [допоміжної статистики](https://uk.wikipedia.org/w/index.php?title=Допоміжна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Ancillary_statistic), [лінійного розрізнювача Фішера](https://uk.wikipedia.org/wiki/Лінійний_дискримінантний_аналіз) та [інформації за Фішером](https://uk.wikipedia.org/wiki/Інформація_за_Фішером).[[31]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-31) У своїй книзі 1930 року [«Генетична теорія природного добору](https://uk.wikipedia.org/w/index.php?title=Генетична_теорія_природного_добору_(книга)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/The_Genetical_Theory_of_Natural_Selection)» він застосував статистику до різних [біологічних](https://uk.wikipedia.org/wiki/Біологія) понять, таких як [принцип Фішера](https://uk.wikipedia.org/wiki/Принцип_Фішера)[[32]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Edwards98-32) (що його [Е. В. Ф. Едвардс](https://uk.wikipedia.org/w/index.php?title=Е._В._Ф._Едвардс&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/A._W._F._Edwards) назвав «імовірно найвизначнішим аргументом в [еволюційній біології»](https://uk.wikipedia.org/wiki/Еволюційна_біологія)) та [фішерова неконтрольованість](https://uk.wikipedia.org/wiki/Фішерова_неконтрольованість),[[33]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-fisher15-33)[[34]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-fisher30-34)[[35]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-pers00-35)[[36]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-ander94-36)[[37]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-ander06-37)[[38]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-gayon10-38) поняття в [статевім доборі](https://uk.wikipedia.org/wiki/Статевий_добір) про афект неконтрольованого зворотного зв'язку, що виникає в [еволюції](https://uk.wikipedia.org/wiki/Еволюція).

Заключна хвиля, яка переважно бачила вдосконалення та розширення попередніх розробок, виникла в результаті спільної праці [Еґона Пірсона](https://uk.wikipedia.org/wiki/Еґон_Пірсон) та [Єжи Неймана](https://uk.wikipedia.org/wiki/Єжи_Нейман) в 1930-х роках. Вони ввели поняття помилки [«II роду»](https://uk.wikipedia.org/wiki/Помилка_другого_роду), [статистичної потужності критерію](https://uk.wikipedia.org/wiki/Статистична_потужність) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *power of a test*), та [довірчих інтервалів](https://uk.wikipedia.org/wiki/Довірчий_інтервал). Єжи Нейман 1934 року показав, що стратифіковане випадкове вибирання було в цілому кращим методом оцінювання, аніж вибирання цілеспрямоване (квотне).[[39]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-39)

Сьогодні статистичні методи застосовують в усіх областях, які передбачають ухвалювання рішень, щоби отримувати точні висновки з консолідованого масиву даних для ухвалювання рішень в умовах невизначеності на основі статистичної методології. Використання сучасних [комп'ютерів](https://uk.wikipedia.org/wiki/Комп'ютер) форсувало великомасштабні статистичні обчислення й також уможливило нові методи, що є недоцільними для виконання вручну. Статистика продовжує бути областю активних досліджень, наприклад, щодо проблеми аналізу [великих даних](https://uk.wikipedia.org/wiki/Великі_дані).[[40]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-40)

## Статистичні дані

*Докладніше:* [*Статистичні дані*](https://uk.wikipedia.org/wiki/Статистичні_дані)

### Збирання даних

#### Вибирання

Коли зібрати дані повного перепису неможливо, статистики збирають вибіркові дані, розробляючи особливі [плани експериментів](https://uk.wikipedia.org/wiki/Планування_експериментів) та [вибірки для обстеження](https://uk.wikipedia.org/w/index.php?title=Вибирання_для_обстеження&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Survey_sampling). Статистика сама по собі також пропонує інструменти для передбачування та прогнозування за допомогою [статистичних моделей](https://uk.wikipedia.org/wiki/Статистична_модель). Ідея робити висновки на основі вибіркових даних виникла близько середини 1600-х у зв'язку з оцінюванням чисельності населення та розробки предтеч страхування життя.[[41]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-41)

Щоби використовувати вибірку як взірець для всієї сукупності, важливо, щоби вона справді представляла генеральну сукупність. Репрезентативне [вибирання](https://uk.wikipedia.org/wiki/Вибирання_(статистика)) забезпечує можливість безпечного розширення висновків та рішень з цієї вибірки на сукупність в цілому. Основна проблема полягає у визначені міри, до якої обрана вибірка є насправді репрезентативною. Статистика пропонує методи для оцінювання та виправляння будь-яких упереджень у вибірці та процедурах збирання даних. Також існують методи планування експериментів для таких експериментів, що можуть зменшувати ці проблеми на початку дослідження, підсилюючи його здатність розпізнавати істину стосовно генеральної сукупності.

Теорія вибирання є частиною [математичної дисципліни](https://uk.wikipedia.org/wiki/Математика) [теорії ймовірності](https://uk.wikipedia.org/wiki/Теорія_ймовірності). Ймовірність використовують в [математичній статистиці](https://uk.wikipedia.org/wiki/Теорія_статистики), щоби досліджувати [вибіркові розподіли](https://uk.wikipedia.org/wiki/Вибірковий_розподіл) [вибіркових статистик](https://uk.wikipedia.org/wiki/Вибіркова_статистика) та, загальніше, властивості [статистичних процедур](https://uk.wikipedia.org/wiki/Статистична_теорія_рішень). Використання будь-якого статистичного методу є правильним, коли система або сукупність, яку розглядають, задовольняє припущення цього методу. Різниця в поглядах класичної теорії ймовірності та теорії вибирання, грубо, полягає в тім, що теорія ймовірності починає з заданих параметрів генеральної сукупності для [дедуктивного виведення](https://uk.wikipedia.org/wiki/Дедукція) ймовірностей, притаманних вибіркам. Проте статистичне висновування рухається в протилежному напрямку, [індуктивно виводячи](https://uk.wikipedia.org/wiki/Індукція_(логіка)) з вибірок параметри більшої або генеральної сукупності.

#### Експериментальні та спостережні дослідження

Загальною метою статистичного дослідницького проєкту є дослідження [причинності](https://uk.wikipedia.org/wiki/Причинність), й зокрема висновування стосовно впливу змін значень передбачувачів чи [незалежних змінних на залежні змінні](https://uk.wikipedia.org/wiki/Залежна_і_незалежна_змінні). Існує два основні типи причиннісних статистичних досліджень: [експериментальні дослідження](https://uk.wikipedia.org/wiki/Експеримент), та [спостережні дослідження](https://uk.wikipedia.org/w/index.php?title=Спостережне_дослідження&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Observational_study). В обох типах досліджень спостерігають за впливом відмінності в незалежній змінній (або змінних) на поведінку залежної змінної. Різниця між цими двома типами полягає в тім, як фактично здійснюють дослідження. Кожен з них може бути дуже дієвим. Експериментальне дослідження включає вимірювання досліджуваної системи, маніпулювання цією системою, а потім здійснення нових вимірювань з використанням тієї ж процедури, щоби визначити, чи змінило це маніпулювання значення вимірювань. На противагу цьому, спостережне дослідження не містить [експериментального маніпулювання](https://uk.wikipedia.org/wiki/Науковий_контроль). Натомість збирають дані та досліджують кореляції між передбачувачами та відгуками. Й хоч інструменти аналізу даних найкраще працюють на даних з [рандомізованих досліджень](https://uk.wikipedia.org/wiki/Рандомізоване_контрольоване_дослідження), їх також застосовують і до інших типів даних, таких як [природні експерименти](https://uk.wikipedia.org/wiki/Природний_експеримент) та [спостережні дослідження](https://uk.wikipedia.org/w/index.php?title=Спостережне_дослідження&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Observational_study),[[42]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-42) для яких статистик використовуватиме видозмінений, структурованіший метод оцінювання (наприклад, серед багатьох інших, [оцінювання різниці в різницях](https://uk.wikipedia.org/w/index.php?title=Різниця_в_різницях&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Difference_in_differences) та [інструментальні змінні](https://uk.wikipedia.org/wiki/Інструментальні_змінні)), що вироблятиме [слушні оцінювачі](https://uk.wikipedia.org/wiki/Слушний_оцінювач).

##### Експерименти

Основними етапами статистичного експерименту є:

1. Планування дослідження, включно зі знаходженням числа повторювань дослідження, із застосуванням наступної інформації: попередніх оцінок стосовно розміру [ефекту впливу](https://uk.wikipedia.org/w/index.php?title=Середній_ефект_впливу&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Average_treatment_effect), [альтернативних гіпотез](https://uk.wikipedia.org/w/index.php?title=Альтернативна_гіпотеза&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Alternative_hypothesis), та оцінюваної [експериментальної мінливості](https://uk.wikipedia.org/wiki/Експериментальна_похибка). Необхідним є розгляд вибору об'єктів експерименту та етики дослідження. Статистики радять, щоб експерименти порівнювали (щонайменше) один новий вплив зі стандартним впливом або керуванням, щоби уможливити неупереджену оцінку відмінності ефектів впливу.
2. [Планування експериментів](https://uk.wikipedia.org/wiki/Планування_експериментів), із застосуванням [групування](https://uk.wikipedia.org/wiki/Групування_експериментів), щоби знижувати вплив [змішувальних змінних](https://uk.wikipedia.org/wiki/Змішувальна_змінна), та [увипадковлених призначень](https://uk.wikipedia.org/w/index.php?title=Увипадковлене_призначування&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Randomized_assignment) впливів до об'єктів, щоби уможливлювати [неупереджені оцінки](https://uk.wikipedia.org/w/index.php?title=Упередження_оцінювача&action=edit&redlink=1) ефектів впливів та експериментальної похибки. На цьому етапі експериментатори та статистики пишуть [*протокол експерименту*](https://uk.wikipedia.org/wiki/Протокол_(природничі_науки)) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *experimental protocol*), що керуватиме виконанням експерименту, й що визначатиме *первинний аналіз* ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *primary analysis*) експериментальних даних.
3. Виконання експерименту згідно [протоколу експерименту](https://uk.wikipedia.org/wiki/Протокол_(природничі_науки)) та [аналізування даних](https://uk.wikipedia.org/wiki/Дисперсійний_аналіз) згідно протоколу експерименту.
4. Подальше вивчення набору даних у вторинних аналізах з метою висування нових гіпотез для майбутнього вивчення.
5. Документування та представлення результатів дослідження.

Експерименти з людською поведінкою несуть особливі турботи. В знаменитім [Готорнськім дослідженні](https://uk.wikipedia.org/wiki/Готорнське_дослідження) вивчали зміни до робочого середовища на Готорнськім заводі компанії [Western Electric](https://uk.wikipedia.org/wiki/Western_Electric). Дослідників цікавило визначити, чи призведе збільшення освітлення до збільшення продуктивності працівників [конвеєра](https://uk.wikipedia.org/w/index.php?title=Конвеєрне_виробництво&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Assembly_line). Дослідники спочатку виміряли продуктивність заводу, потім змінили освітлення в одній області заводу, й перевірили, чи вплинули ці зміни в освітленні на продуктивність. Виявилося, що продуктивність і справді покращилася (в експериментальних умовах). Проте це дослідження сьогодні сильно критикують через помилки в процедурах експерименту, особливо через брак [контрольної групи](https://uk.wikipedia.org/wiki/Контрольна_група) та [сліпоти](https://uk.wikipedia.org/wiki/Подвійно_сліпе_дослідження). [Готорнський ефект](https://uk.wikipedia.org/wiki/Готорнський_ефект) стосується виявлення того, що результат (в цьому випадку — продуктивність працівників) змінився через саме спостереження. Піддослідні в Готорнськім дослідженні стали продуктивнішими не через зміну освітлення, а через те, що за ними спостерігали.[[43]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-pmid17608932-43)

##### Спостережне дослідження

Прикладом спостережного дослідження є таке, що вивчає пов'язаність паління та раку легенів. Цей тип дослідження зазвичай використовує опитування для збирання спостережень про цільову область, і потім виконує статистичний аналіз. В цьому випадку дослідники збирали би спостереження як про курців, так і про не курців, певно, шляхом [когортного дослідження](https://uk.wikipedia.org/wiki/Когортне_дослідження), а потім дивилися би на число випадків раку легенів у кожній з груп.[[44]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-44) Іншим типом спостережного дослідження є [дослідження «випадок—контроль»](https://uk.wikipedia.org/wiki/Дослідження_«випадок—контроль»), в якому запрошують взяти участь людей з та без цільового результату (наприклад, раку легенів), і збирають їхні історії піддавання впливові.

### Типи даних

*Детальніші відомості з цієї теми ви можете знайти в статті* [***Типи статистичних даних***](https://uk.wikipedia.org/w/index.php?title=Типи_статистичних_даних&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Statistical_data_type) ***та*** [***Шкала***](https://uk.wikipedia.org/wiki/Шкала)*.*

Існувало чимало спроб виробити таксономію [шкал вимірювання](https://uk.wikipedia.org/wiki/Шкала). Психолог [Стенлі Сміт Стівенс](https://uk.wikipedia.org/w/index.php?title=Стенлі_Сміт_Стівенс&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Stanley_Smith_Stevens) визначив номінальну ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *nominal*), порядкову ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *ordinal*), інтервальну ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *interval*) шкали, та шкалу відношень ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *ratio scale*). Номінальні вимірювання не мають змістовного порядку ранжування їхніх значень, й дозволяють будь-яке перетворення один-в-одного (ін'єктивне). Порядкові вимірювання мають неточні відмінності між послідовними значеннями, але мають змістовний порядок цих значень, й дозволяють будь-яке перетворення зі збереженням порядку. Інтервальні вимірювання мають визначені змістовні відстані між вимірюваннями, але нульове значення є довільним (як у випадках вимірювань [довготи](https://uk.wikipedia.org/wiki/Довгота) та [температури](https://uk.wikipedia.org/wiki/Температура) в [градусах Цельсія](https://uk.wikipedia.org/wiki/Градус_Цельсія) та [Фаренгейта](https://uk.wikipedia.org/wiki/Шкала_Фаренгейта)), й дозволяють будь-яке лінійне перетворення. Вимірювання відношень мають визначені змістовні як нульове значення, так і відстані між вимірюваннями, й дозволяють будь-яке перетворення масштабування.

Оскільки змінні, що відповідають лише номінальним та порядковим вимірюванням, раціонально виміряти числами неможливо, іноді їх об'єднують як [категорійні змінні](https://uk.wikipedia.org/wiki/Категорійна_змінна), тоді як відносні та інтервальні вимірювання об'єднують як [кількісні змінні](https://uk.wikipedia.org/wiki/Кількісна_змінна), що можуть бути або [дискретними](https://uk.wikipedia.org/wiki/Дискретний_розподіл_імовірності), або [неперервними](https://uk.wikipedia.org/wiki/Неперервний_розподіл_імовірності), в силу своєї числової природи. Таке розмежування часто може бути нестрого співвідносним з [типом даних](https://uk.wikipedia.org/wiki/Тип_даних) в інформатиці, оскільки дихотомні категорійні змінні може бути представлено [логічним типом даних](https://uk.wikipedia.org/wiki/Логічний_тип_даних), багатозначні категорійні змінні — довільно призначуваними [цілими числами](https://uk.wikipedia.org/wiki/Цілі_числа) в [цілочисловім типі даних](https://uk.wikipedia.org/wiki/Цілочисловий_тип_даних), а неперервні змінні — [дійснозначним типом даних](https://uk.wikipedia.org/w/index.php?title=Дійснозначний_тип_даних&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Real_data_type) із застосуванням обчислень [з рухомою комою](https://uk.wikipedia.org/wiki/Число_з_рухомою_комою). Але відображення типів даних інформатики на типи статистичних даних залежить від того, яку категоризацію останніх втілюють.

Було запропоновано й інші категоризації. Наприклад, Мостеллер та Тьюкі (1997)[[45]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-45) розрізнювали ступені ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *grades*), ранги ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *ranks*), зліченні дроби ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *counted fractions*), кількості ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *counts*), суми ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *amounts*) та баланси ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *balances*). Нелдер (1990)[[46]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-46) описав неперервні кількості ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *continuous counts*), неперервні відношення ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *continuous ratios*), відношення кількостей ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *count ratios*), та категорійні види даних. Див. також Крісмана (1998),[[47]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-47) ван ден Берґа (1991).[[48]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-48)

Питання доречності чи недоречності застосування різних видів статистичних методів до даних, отриманих з різних видів процедур вимірювання, ускладнюється питаннями перетворювання змінних та точною інтерпретацією досліджуваних питань. «Взаємозв'язок між даними та тим, що вони описують, просто відображає той факт, що певні види статистичних висловлень можуть мати значення істинності, що не є інваріантними за деяких перетворень. Чи є доцільним розглядати певне перетворення, чи ні, залежить від питання, на яке намагаються відповісти.»[[49]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-49):82

## Статистичні методи

### Описова статистика

*Докладніше:* [*Описова статистика*](https://uk.wikipedia.org/wiki/Описова_статистика)

**Описо́ві стати́стики** (в сенсі [злічуваного іменника](https://uk.wikipedia.org/w/index.php?title=Злічуваний_іменник&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Count_noun) [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *descriptive statistic* в однині) — це [зведені статистики](https://uk.wikipedia.org/wiki/Зведена_статистика), які кількісно описують або узагальнюють ознаки сукупності [інформації](https://uk.wikipedia.org/wiki/Інформація),[[50]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-50) в той час як **опис́ова стати́стика** в сенсі [незлічуваного іменника](https://uk.wikipedia.org/wiki/Незлічуваний_іменник) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *descriptive statistics*) — це процес використання та аналізу цих статистик. Описова статистика відрізняється від [висновувальної статистики](https://uk.wikipedia.org/wiki/Висновувальна_статистика)[[51]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-51) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *inferential statistics*, або індуктивної статистики, [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *inductive statistics*) тим, що описова статистика має на меті узагальнювання [вибірки](https://uk.wikipedia.org/wiki/Вибірка), а не використання цих даних, щоб дізнатися щось про [генеральну сукупність](https://uk.wikipedia.org/wiki/Статистична_сукупність), яку, як вважають, ця вибірка даних представляє.

### Індуктивна статистика

*Докладніше:* [*Статистичне висновування*](https://uk.wikipedia.org/wiki/Статистичне_висновування)

**Статисти́чне висно́вування** ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistical inference*) — це процес використання [аналізу даних](https://uk.wikipedia.org/wiki/Аналіз_даних) для встановлення властивостей [розподілу ймовірності](https://uk.wikipedia.org/wiki/Розподіл_імовірності), що лежить в їх основі.[[52]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Oxford-52) Висновувальний статистичний аналіз робить висновки про властивості [генеральної сукупності](https://uk.wikipedia.org/wiki/Генеральна_сукупність), наприклад, шляхом перевіряння гіпотез та отримування оцінок. Він виходить з припущення, що спостережувані дані є [вибіркою](https://uk.wikipedia.org/wiki/Відбір_вибірки_(статистика)) з більшої сукупності. Індуктивну статистику можливо протиставляти [описовій статистиці](https://uk.wikipedia.org/wiki/Описова_статистика). Описова статистика цікавиться виключно властивостями спостережуваних даних, і не спирається на припущення, що ці дані походять із більшої сукупності.

#### Термінологія та теорія індуктивної статистики

##### Статистики, статистичні оцінки, та центральні величини

Розгляньмо [незалежні однаково розподілені (н. о. р.) випадкові змінні](https://uk.wikipedia.org/wiki/Незалежні_однаково_розподілені_випадкові_величини) із заданим [розподілом ймовірності](https://uk.wikipedia.org/wiki/Розподіл_імовірності): стандартне [статистичне висновування](https://uk.wikipedia.org/wiki/Статистичне_висновування) та [теорія оцінювання](https://uk.wikipedia.org/wiki/Теорія_оцінювання) визначає [випадкову вибірку](https://uk.wikipedia.org/wiki/Випадкова_вибірка) як [випадковий вектор](https://uk.wikipedia.org/wiki/Випадковий_вектор), заданий [стовпчиковим вектором](https://uk.wikipedia.org/wiki/Стовпчиковий_вектор) цих н. о. р. змінних.[[53]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Piazza-53) Досліджувану [генеральну сукупність](https://uk.wikipedia.org/wiki/Генеральна_сукупність) описують розподілом ймовірності, що може мати невідомі параметри.

[Статистика](https://uk.wikipedia.org/wiki/Статистика_(математика)) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistic*) — це випадкова змінна, що є функцією випадкової вибірки, але *не функцією невідомих параметрів*. Розподіл імовірності цієї статистики, проте, невідомі параметри мати може.

Розгляньмо тепер функцію невідомого параметра: [статистична оцінка](https://uk.wikipedia.org/wiki/Статистична_оцінка) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *estimator*) — це статистика, яку використовують для оцінювання цієї функції. До широко вживаних статистичних оцінок належать [вибіркове середнє](https://uk.wikipedia.org/wiki/Вибіркове_середнє), незміщена [дисперсія вибірки](https://uk.wikipedia.org/wiki/Дисперсія_вибірки) та [коваріація вибірки](https://uk.wikipedia.org/w/index.php?title=Коваріація_вибірки&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Sample_covariance).

Випадкову змінну, що є функцією випадкової вибірки та невідомого параметру, але чий розподіл імовірності *не залежить від невідомого параметру*, називають [центральною величиною](https://uk.wikipedia.org/w/index.php?title=Центральна_величина&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Pivotal_quantity) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *pivotal quantity, pivot*). До широко вживаних центральних величин належать [*z*-оцінка](https://uk.wikipedia.org/wiki/Z-оцінка), [статистика хі-квадрат](https://uk.wikipedia.org/wiki/Розподіл_хі-квадрат" \l "Застосування) та *[t](https://uk.wikipedia.org/wiki/T-розподіл_Стьюдента" \l "Як_розподіл_Стьюдента_виникає_з_вибірки)*-величина Стьюдента.

Серед двох оцінок заданого параметру [ефективнішою](https://uk.wikipedia.org/wiki/Ефективна_оцінка) вважають ту, що має нижчу [середньоквадратичну похибку](https://uk.wikipedia.org/wiki/Середньоквадратична_похибка). Крім того, оцінку називають [незміщеною](https://uk.wikipedia.org/wiki/Незміщена_оцінка) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *unbiased*), якщо її [математичне сподівання](https://uk.wikipedia.org/wiki/Математичне_сподівання) дорівнює істинному значенню оцінюваного невідомого параметра, й асимптотично незміщеною, якщо її математичне сподівання збігається до [границі](https://uk.wikipedia.org/wiki/Границя) істинного значення такого параметра.

До інших бажаних властивостей статистичних оцінок належать: [рівномірно незміщені оцінки з мінімальною дисперсією](https://uk.wikipedia.org/w/index.php?title=Рівномірно_незміщена_оцінка_з_мінімальною_дисперсією&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Uniformly_minimum-variance_unbiased_estimator) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *UMVUE*), що мають найнижчу дисперсію для всіх можливих значень оцінюваного параметра (це зазвичай є легшою властивістю для перевірки, ніж ефективність), та [слушні оцінки](https://uk.wikipedia.org/wiki/Слушна_оцінка) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *consistent estimators*), що [збігаються за ймовірністю](https://uk.wikipedia.org/wiki/Збіжність_за_ймовірністю) до істинного значення такого параметра.

Це все ще залишає відкритим питання, як отримувати статистичні оцінки в заданій ситуації та виконувати обчислення, було запропоновано декілька методів: [метод моментів](https://uk.wikipedia.org/wiki/Метод_моментів), [метод максимальної правдоподібності](https://uk.wikipedia.org/wiki/Метод_максимальної_правдоподібності), [метод найменших квадратів](https://uk.wikipedia.org/wiki/Метод_найменших_квадратів), та новіший метод [оцінних рівнянь](https://uk.wikipedia.org/w/index.php?title=Оцінні_рівняння&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Estimating_equations).

##### Нульова гіпотеза та альтернативна гіпотеза

Інтерпретування статистичної інформації часто може включати розробку [нульової гіпотези](https://uk.wikipedia.org/wiki/Нульова_гіпотеза) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *null hypothesis*), яка зазвичай (але не обов'язково) полягає у відсутності взаємозв'язку серед змінних, або що зміни з часом не відбуваються.[[54]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-54)[[55]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-55)

Найкращою ілюстрацією для новачка є утруднення, з яким зіткнувся кримінальний процес в суді присяжних. Нульова гіпотеза, *H*0, стверджує, що відповідач є невинним, тоді як альтернативна гіпотеза, *H*1, стверджує, що відповідач є винним. Висувається звинувачення через підозру в винності. *H*0 (статус кво) протистоїть *H*1, й підтримується, поки *H*1 не стане підтримано доказами «поза розумним сумнівом». Проте «нездатність відхилити *H*0» в цьому випадку означає не невинність, а лише те, що докази були недостатніми для засудження. Тож присяжні не обов'язково *приймають* *H*0, їм *не вдається відхилити* *H*0. І хоч «довести» нульову гіпотезу неможливо, її можливо перевірити на те, наскільки вона є близькою до істини, через [статистичну потужність](https://uk.wikipedia.org/wiki/Статистична_потужність), яка робить перевірку на [помилки другого роду](https://uk.wikipedia.org/wiki/Помилка_другого_роду).

Те, що [статистики](https://uk.wikipedia.org/wiki/Статистик) називають [альтернативною гіпотезою](https://uk.wikipedia.org/w/index.php?title=Альтернативна_гіпотеза&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Alternative_hypothesis), — це просто гіпотеза, що суперечить [нульовій гіпотезі](https://uk.wikipedia.org/wiki/Нульова_гіпотеза).

##### Похибка

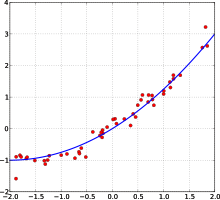
Працюючи від [нульової гіпотези](https://uk.wikipedia.org/wiki/Нульова_гіпотеза), розпізнаю́ть два основні види помилок:

* [Помилки I роду](https://uk.wikipedia.org/wiki/Помилки_I_роду) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *type I errors*), коли нульову гіпотезу хибно відхиляють, що дає «хибно позитивне» ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *false positive*)
* [Помилки II роду](https://uk.wikipedia.org/wiki/Помилки_II_роду) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *type II errors*), коли нульову гіпотезу не вдається відхилити, й справжня відмінність між сукупностями втрачається, що дає «хибно негативне» ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *false negative*)

[Стандартне відхилення](https://uk.wikipedia.org/wiki/Стандартне_відхилення) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *standard deviation*) вказує на те, наскільки окремі спостереження в вибірці відрізняються від центрального значення, такого як середнє за вибіркою або середнє за генеральною сукупністю, тоді як [стандартна похибка](https://uk.wikipedia.org/wiki/Стандартна_похибка" \l "Стандартна_похибка_середнього) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *standard error*) вказує на оцінку різниці між середніми за вибіркою та середнім за генеральною сукупністю.

[Статистична похибка](https://uk.wikipedia.org/wiki/Похибки_та_залишки_у_статистиці" \l "Введення) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistical error*) — це величина, на яку спостереження відрізняються від їхнього [математичного сподівання](https://uk.wikipedia.org/wiki/Математичне_сподівання), [залишок](https://uk.wikipedia.org/wiki/Похибки_та_залишки_у_статистиці" \l "Введення) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *residual*) — це величина, на яку спостереження відрізняються від значення, якого набуває статистична оцінка очікуваного значення на заданому зразкові (яку також називають передбаченням).

[Середньоквадратичну похибку](https://uk.wikipedia.org/wiki/Середньоквадратична_похибка) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *mean squared error*) використовують для отримування [ефективних оцінок](https://uk.wikipedia.org/wiki/Ефективна_оцінка), широко вживаного класу статистичних оцінок. [Коренева середньоквадратична похибка](https://uk.wikipedia.org/w/index.php?title=Коренева_середньоквадратична_похибка&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Root_mean_square_error) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *root mean square error*) є просто квадратним коренем середньоквадратичної похибки.

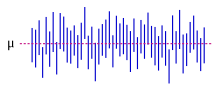
[](https://uk.wikipedia.org/wiki/Файл:Linear_least_squares(2).svg)

Багато статистичних методів прагнуть мінімізувати [залишкову суму квадратів](https://uk.wikipedia.org/w/index.php?title=Залишкова_сума_квадратів&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Residual_sum_of_squares) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *residual sum of squares*), і їх називають [методами найменших квадратів](https://uk.wikipedia.org/wiki/Метод_найменших_квадратів) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *methods of least squares*), на противагу до [методів найменших модулів](https://uk.wikipedia.org/w/index.php?title=Метод_найменших_модулів&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Least_absolute_deviations) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *least absolute deviations*). Останні надають однакової ваги як маленьким, так і великим похибкам, тоді як перші надають великим похибкам більшої ваги. Також, залишкова сума квадратів є [диференційовною](https://uk.wikipedia.org/wiki/Диференційовна_функція), що забезпечує зручну властивість для виконання [регресії](https://uk.wikipedia.org/wiki/Регресійний_аналіз). Найменші квадрати в застосуванні до [лінійної регресії](https://uk.wikipedia.org/wiki/Лінійна_регресія) називають [звичайним методом найменших квадратів](https://uk.wikipedia.org/w/index.php?title=Звичайний_метод_найменших_квадратів&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Ordinary_least_squares) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *ordinary least squares method*), а найменші квадрати в застосуванні до [нелінійної регресії](https://uk.wikipedia.org/w/index.php?title=Нелінійна_регресія&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Nonlinear_regression) називають [нелінійним методом найменших квадратів](https://uk.wikipedia.org/w/index.php?title=Нелінійний_метод_найменших_квадратів&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Non-linear_least_squares) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *non-linear least squares*). Також, в лінійній регресійній моделі недетерміновану частину моделі називають членом похибки ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *error term*), збуренням ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *disturbance*), або просто шумом ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *noise*). Як лінійну, так і нелінійну регресію розглядають у [поліноміальнім методі найменших квадратів](https://uk.wikipedia.org/wiki/Поліноміальний_метод_найменших_квадратів) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *polynomial least squares*), що також описує дисперсію в передбаченні залежної змінної (вісь *y*) як функцію від незалежної змінної (вісь *x*) та відхилень (похибок, шуму, збурення) відносно оцінюваної (допасовуваної) кривої.

Процеси вимірювання, що породжують статистичні дані, також є схильними до похибок. Багато з цих похибок класифікують як [випадкові](https://uk.wikipedia.org/wiki/Випадкова_похибка) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *random*, шум) та [систематичні](https://uk.wikipedia.org/wiki/Систематична_похибка) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *systematic*, [зсув](https://uk.wikipedia.org/wiki/Зсув_оцінки)), але важливими можуть бути й інші типи похибок (наприклад, промахи, такі як коли аналітик повідомляє неправильні одиниці вимірювання). Існування [пропущених даних](https://uk.wikipedia.org/w/index.php?title=Пропущені_дані&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Missing_data) або [цензурування](https://uk.wikipedia.org/w/index.php?title=Цензурування_(статистика)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Censoring_(statistics)) може призводити до [упереджених оцінок](https://uk.wikipedia.org/wiki/Упередження_(статистика)), й для подолання цих проблем було розроблено спеціальні методики.[[56]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-56)

##### Інтервальне оцінювання

*Детальніші відомості з цієї теми ви можете знайти в статті* [***Інтервальне оцінювання***](https://uk.wikipedia.org/w/index.php?title=Інтервальне_оцінювання&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Interval_estimation)*.*

[](https://uk.wikipedia.org/wiki/Файл:NYW-confidence-interval.svg)[Довірчі інтервали](https://uk.wikipedia.org/wiki/Довірчий_інтервал): червона лінія є істинним значенням середнього в цьому прикладі, а сині лінії є випадковими довірчими інтервалами для 100 реалізацій.

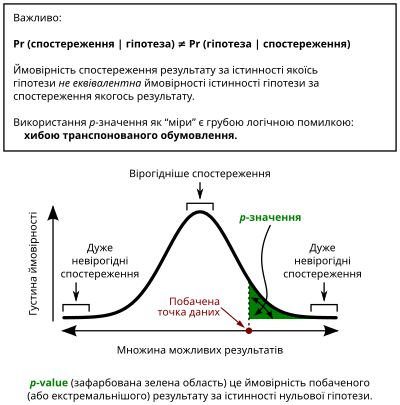
Більшість досліджень вибирають лише частину генеральної сукупності, тож результати не представляють всю генеральну сукупність вповні. Будь-які оцінки, отримані з такої вибірки, лише наближують значення генеральної сукупності. [Довірчі інтервали](https://uk.wikipedia.org/wiki/Довірчий_інтервал) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *confidence intervals*) дозволяють статистикам виражати те, наскільки близько вибіркова оцінка відповідає істинному значенню для всієї генеральної сукупності. Часто їх виражають як 95-відсоткові довірчі інтервали. Формально, 95 %-вий довірчий інтервал для значення є діапазоном, який, якщо вибирання та аналіз повторювати за таких же умов (отримуючи відмінний набір даних), включатиме істинне значення (генеральної сукупності) в 95 % всіх можливих випадків. Це *не* означає, що ймовірність перебування істинного значення в цьому довірчому інтервалі становить 95 %. З [частотницької](https://uk.wikipedia.org/wiki/Частотницьке_висновування) точки зору таке твердження не має сенсу, оскільки істинне значення не є [випадковою змінною](https://uk.wikipedia.org/wiki/Випадкова_змінна). Істинне значення або перебуває в даному інтервалі, або ні. Проте, істинним є те, що до того, як буде вибрано якісь дані, і за заданого плану побудови довірчого інтервалу, ймовірність того, що інтервал, який ще належить обчислити, покриватиме істинне значення, становить 95 %: в цей момент межі інтервалу є [випадковими змінними](https://uk.wikipedia.org/wiki/Випадкова_змінна), які ще належить проспостерігати. Одним із підходів, що видає інтервал, який можливо інтерпретувати як такий, що має задану ймовірність вміщування істинного значення, є застосування [ймовірних інтервалів](https://uk.wikipedia.org/wiki/Імовірний_інтервал) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *credible intervals*) з [баєсової статистики](https://uk.wikipedia.org/wiki/Баєсова_статистика): цей підхід залежить від відмінного способу [інтерпретування того, що мається на увазі під «імовірністю»](https://uk.wikipedia.org/wiki/Інтерпретації_ймовірності), а саме, [баєсової ймовірності](https://uk.wikipedia.org/wiki/Баєсова_ймовірність).

Довірчі інтервали, в принципі, можуть бути симетричними та асиметричними. Інтервал може бути асиметричним, бо він працює як нижня та верхня межі для параметру (лівобічний та правобічний інтервали), але він також може бути асиметричним, оскільки цей двобічний інтервал будують із порушенням симетрії навколо оцінки. Іноді межі асимптотичного інтервалу досягають асимптотично, й використовують їх для наближення істинних меж.

##### Значущість

*Докладніше:* [*Статистична значущість*](https://uk.wikipedia.org/wiki/Статистична_значущість)

Статистика рідко дає на аналізоване питання просту відповідь на кшталт Так/Ні. Інтерпретація часто зводиться до рівня статистичної значущості ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistical significance*), застосовуваного до чисел, і часто посилається на ймовірність значення, що точно відкидає нульову гіпотезу (яку іноді називають [*p*-значенням](https://uk.wikipedia.org/wiki/P-значення)).

[](https://uk.wikipedia.org/wiki/Файл:P-value_in_statistical_significance_testing_uk.svg)[статистики критерію](https://uk.wikipedia.org/wiki/Статистичний_критерій), [критична область](https://uk.wikipedia.org/wiki/Перевірка_статистичних_гіпотез" \l "Визначення_термінів) є множиною значень праворуч від спостережуваної точки даних (спостережуваного значення статистичного критерію), а [*p*-значення](https://uk.wikipedia.org/wiki/P-значення) представлено зеленою площею.

Стандартним підходом[[53]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Piazza-53) є перевіряти нульову гіпотезу відносно альтернативної гіпотези. [Критична область](https://uk.wikipedia.org/wiki/Перевірка_статистичних_гіпотез" \l "Визначення_термінів) є множиною значень оцінювача, які ведуть до спростування нульової гіпотези. Ймовірність помилки I роду є відтак ймовірністю того, що оцінювач лежить у критичній області за умови, що нульова гіпотеза є істинною ([статистична значущість](https://uk.wikipedia.org/wiki/Статистична_значущість)), а ймовірність помилки II роду є ймовірністю того, що оцінювач не належить до критичної області за умови, що істинною є альтернативна гіпотеза. [Статистична потужність](https://uk.wikipedia.org/wiki/Статистична_потужність) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistical power*) критерію є ймовірністю того, що він належним чином відхиляє нульову гіпотезу, коли ця нульова гіпотеза є хибною.

Посилання на статистичну значущість не обов'язково означає, що загальний результат є значущим в термінах реального світу. Наприклад, у великому дослідженні лікарського засобу може бути показано, що цей медикамент має статистично значущий, але дуже маленький сприятливий ефект, такий, що він навряд чи може помітно допомагати пацієнтові.

І хоча в принципі прийнятний рівень [статистичної значущості](https://uk.wikipedia.org/wiki/Статистична_значущість) може бути предметом обговорення, [*p*-значення](https://uk.wikipedia.org/wiki/P-значення) є найменшим рівнем значущості, який дозволяє критерію відхиляти нульову гіпотезу. Цей критерій є логічно рівнозначним твердженню, що *p*-значення є ймовірністю спостерігання результату, щонайменше настільки ж екстремального, як і [статистика критерію](https://uk.wikipedia.org/wiki/Статистичний_критерій), за умови, що нульова гіпотеза є істинною. Таким чином, що меншим є *p*-значення, то нижчою є ймовірність трапляння помилки I роду.

Деякі проблеми, зазвичай пов'язувані з цією системою (див. [критику перевірки статистичних гіпотез](https://uk.wikipedia.org/w/index.php?title=Критика_перевірки_статистичних_гіпотез&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Criticism_of_statistical_hypothesis_testing)):

* Відмінність, яка має високу статистичну значущість, може все одно не мати практичної значущості, але можливо належно формулювати критерії, щоби враховувати це. Одна з відповідей передбачає вихід за межі повідомляння лише [рівня значущості](https://uk.wikipedia.org/wiki/Рівень_значущості), шляхом включання [*p*-значення](https://uk.wikipedia.org/wiki/P-значення) при звітуванні про відхилення чи прийняття гіпотези. Проте, *p*-значення не показує [розміру](https://uk.wikipedia.org/w/index.php?title=Розмір_ефекту&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Effect_size) чи важливості спостережуваного ефекту, а також може створювати враження перебільшеної важливості незначних відмінностей у великих дослідженнях. Кращим і все поширенішим підходом є повідомляти [довірчі інтервали](https://uk.wikipedia.org/wiki/Довірчий_інтервал). Незважаючи на те, що їх отримують із тих же обчислень, що й статистичні критерії гіпотез та *p*-значення, вони описують розмір як самого ефекту, так і невизначеності, що його оточує.
* Хиба транспонованого обумовлення, вона ж [помилка прокурора](https://uk.wikipedia.org/wiki/Помилка_прокурора): критика виникає через те, що цей підхід перевірки гіпотез змушує віддавати перевагу одній з гіпотез ([нульовій](https://uk.wikipedia.org/wiki/Нульова_гіпотеза)), оскільки оцінюють ймовірність спостережуваного результату за умови нульової гіпотези, а не ймовірність нульової гіпотези за спостереженого результату. Альтернативу цьому підходові запропоновано [баєсовим висновуванням](https://uk.wikipedia.org/wiki/Баєсове_висновування), хоча воно вимагає встановлювання [апріорної ймовірності](https://uk.wikipedia.org/wiki/Апріорна_ймовірність).[[57]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Ioannidis2005-57)
* Відхилення нульової гіпотези не доводить автоматично альтернативну гіпотезу.
* Як і все в [індуктивній статистиці](https://uk.wikipedia.org/wiki/Індуктивна_статистика), вона покладається на розмір вибірки, й відтак за [важких хвостів](https://uk.wikipedia.org/w/index.php?title=Важкий_хвіст&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Fat_tail) *p*-значення може бути обчислювано сильно помилково.[[*прояснити*](https://uk.wikipedia.org/wiki/Шаблон:Прояснити/документація)]

##### Приклади

Деякі відомі статистичні [критерії](https://uk.wikipedia.org/wiki/Статистична_перевірка_гіпотез) та процедури:

* [Дисперсійний аналіз](https://uk.wikipedia.org/wiki/Дисперсійний_аналіз) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *ANOVA*)
* [Критерій хі-квадрат](https://uk.wikipedia.org/wiki/Критерій_хі-квадрат)
* [Кореляція](https://uk.wikipedia.org/wiki/Кореляція)
* [Факторний аналіз](https://uk.wikipedia.org/wiki/Факторний_аналіз)
* [*U* Манна—Уітні](https://uk.wikipedia.org/wiki/U-критерій_Манна-Уітні)
* [Зважене середньоквадратичне відхилення](https://uk.wikipedia.org/w/index.php?title=Зважене_середньоквадратичне_відхилення&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Mean_square_weighted_deviation) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *MSWD*)
* [Коефіцієнт кореляції моменту добутку Пірсона](https://uk.wikipedia.org/wiki/Коефіцієнт_кореляції_моменту_добутку_Пірсона)
* [Регресійний аналіз](https://uk.wikipedia.org/wiki/Регресійний_аналіз)
* [Коефіцієнт кореляції рангу Спірмена](https://uk.wikipedia.org/wiki/Коефіцієнт_кореляції_рангу_Спірмена)
* [*t*-критерій Стьюдента](https://uk.wikipedia.org/wiki/T-критерій_Стьюдента)
* [Аналіз часових рядів](https://uk.wikipedia.org/wiki/Аналіз_часових_рядів)
* [Спільний аналіз](https://uk.wikipedia.org/w/index.php?title=Спільний_аналіз&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Conjoint_Analysis)[[58]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-58)

### Розвідувальний аналіз даних

*Докладніше:* [*Розвідувальний аналіз даних*](https://uk.wikipedia.org/wiki/Розвідувальний_аналіз_даних)

**Розві́дувальний ана́ліз да́них** (**РАД**, [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *exploratory data analysis, EDA*) — це один з підходів до [аналізу](https://uk.wikipedia.org/wiki/Аналіз_даних) [наборів даних](https://uk.wikipedia.org/wiki/Набір_даних) для узагальнювання їхніх основних характеристик, часто за допомогою візуальних методів. [Статистичну модель](https://uk.wikipedia.org/wiki/Статистична_модель) можуть використовувати чи ні, але в першу чергу РАД призначено, щоби побачити, що дані можуть сказати нам за межами формальної задачі моделювання та перевірки гіпотез.

## Неналежне застосування

*Детальніші відомості з цієї теми ви можете знайти в статті* [***Неналежне застосування статистики***](https://uk.wikipedia.org/wiki/Неналежне_застосування_статистики)*.*

[Неналежне застосування статистики](https://uk.wikipedia.org/wiki/Неналежне_застосування_статистики) може спричинювати тонкі, але серйозні помилки в описі та інтерпретації — тонкі в тому сенсі, що навіть досвідчені фахівці можуть робити такі помилки, а серйозні в тому сенсі, що вони можуть призводити до руйнівних помилок в ухвалюванні рішень. Від належного застосування статистики залежать, наприклад, соціальна політика, медична практика та надійність таких споруд, як мости.

Навіть коли статистичні методики застосовують коректно, їхні результати можуть бути складними для інтерпретування для тих, кому бракує досвіду. [Статистична значущість](https://uk.wikipedia.org/wiki/Статистична_значущість) тенденції в даних, яка вимірює ступінь, до якого тенденцію може бути спричинено випадковою мінливістю вибірки, може узгоджуватися з інтуїтивним відчуттям її значущості, а може й не узгоджуватися. Набір базових статистичних навичок (та скептицизму), необхідних людям для належної роботи з інформацією у своєму повсякденному житті, називають [статистичною грамотністю](https://uk.wikipedia.org/w/index.php?title=Статистична_грамотність&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Statistical_literacy).

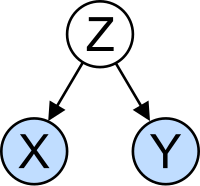
Існує думка, що статистичні знання занадто часто цілеспрямовано [застосовують неналежним чином](https://uk.wikipedia.org/wiki/Неналежне_застосування_статистики), шукаючи шляхи інтерпретувати лише ті дані, що є сприятливими для доповідача.[[59]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Huff-59) Недовіра та нерозуміння статистики пов'язані з цитатою [«Існує три види брехні: брехня, нахабна брехня й статистика»](https://uk.wikipedia.org/wiki/Брехня,_нахабна_брехня_й_статистика). Неналежне застосування статистики може бути як ненавмисним, так і навмисним, й ряд міркувань окреслено в книзі [«Як брехати за допомогою статистики»](https://uk.wikipedia.org/wiki/Як_брехати_за_допомогою_статистики).[[59]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Huff-59) В намаганні пролити світло на належне та неналежне застосування статистики проводять перегляди статистичних методик, використовуваних в певних областях (наприклад, Варн, Лазо, Рамос і Ріттер (2012)).[[60]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-60)

До способів запобігання неналежному застосуванню статистики належать застосування правильних діаграм та запобігання [упередженню](https://uk.wikipedia.org/wiki/Упередження_(статистика)).[[61]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Statistics_in_Archaeology-61) Неналежне застосування може траплятися, коли висновки [переузагальнюють](https://uk.wikipedia.org/wiki/Квапливе_узагальнення), і претендують, що вони є репрезентативними для більшого, ніж вони є насправді, часто через навмисне або ненавмисне не враховування упередженості вибірки.[[62]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Misuse_of_Statistics-62) [Стовпчикові діаграми](https://uk.wikipedia.org/wiki/Стовпчикова_діаграма) є, мабуть, найпростішими діаграмами для застосування та розуміння, й їх можливо робити вручну або за допомогою простих комп'ютерних програм.[[61]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Statistics_in_Archaeology-61) На жаль, більшість людей не дивляться на упередження та помилки, тож вони лишаються непоміченими. Таким чином, люди можуть часто вірити в те, що щось є істинним, навіть якщо воно не є достатньо [репрезентативним](https://uk.wikipedia.org/wiki/Вибирання_(статистика)).[[62]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Misuse_of_Statistics-62) Щоби зібрані зі статистики дані були правдоподібними та точними, здійснювана вибірка мусить бути репрезентативною для генеральної сукупності в цілому.[[63]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Modern_Elementary_Statistics-63) Згідно Гаффа, «Надійність вибірки може бути зруйновано [упередженістю]… дозволяйте собі певну міру скептицизму.»[[64]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-64)

Для допомоги в розумінні статистики Гафф запропонував ряд питань, які слід задавати в кожному випадку:[[59]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-Huff-59)

* Хто так каже? (Чи має він/вона корисливу мету?)
* Звідки він/вона знає? (Чи має він/вона ресурси, щоби знати ці факти?)
* Що упущено? (Чи надали він/вона повну картину?)
* Чи не змінив хтось тему? (Чи пропонують нам він/вона правильну відповідь на не ту задачу?)
* Чи має це сенс? (Чи є його/її висновок логічним та узгодженим з тим, що ми вже знаємо?)

[Стентон Ґлентц](https://uk.wikipedia.org/w/index.php?title=Стентон_Ґлентц&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Stanton_Glantz), американський професор медицини, який викладає статистику студентам медичного профілю, автор ряду курсів та книг зі статистики та колишній редактор [«Journal of the American College of Cardiology](https://uk.wikipedia.org/w/index.php?title=Journal_of_the_American_College_of_Cardiology&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Journal_of_the_American_College_of_Cardiology)», зазначає, що близько 50 % публікацій, які надходять в редакцію журналу, містять статистичні помилки отримання та обробки медичних даних.[[65]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-glanc_gur-65) Основними причинами є недостовірність отриманих даних (неправильно поставлені експерименти або ж взяті непрезентабельні дані), а також незнання та неправильне застосування статистичних методів. Також він зазначає, що часто дослідник та особи, причетні до експерименту, можуть підсвідомо видавати бажане за дійсне. Причому ненавмисне підтасування даних може відбуватися як на етапі постановки експерименту, збору даних, так і на етапі аналізу даних. Виходом є максимальне врахування та усунення сторонніх чинників, які можуть вплинути на процес експерименту та на аналіз даних. Він пропонує якомога ширше використовувати [«сліпий метод»](https://uk.wikipedia.org/wiki/Сліпий_експеримент) чи навіть [«подвійно сліпий метод»](https://uk.wikipedia.org/wiki/Подвійно_сліпе_дослідження), коли ані піддослідні, ані дослідники (чи помічники дослідників) достеменно не знають, що на якій групі хворих досліджується, і навіть аналіз даних бажано щоб робила особа, незацікавлена у некоректній інтерпретації даних, чи, ще краще, якщо вона буде необізнаною в конкретних деталях експерименту. В будь-якому разі, в постановці, зборі та аналізі даних повинні брати участь особи, які добре володіють прикладними статистичними методами.

[](https://uk.wikipedia.org/wiki/Файл:Simple_Confounding_Case.svg)[змішувальної змінної](https://uk.wikipedia.org/wiki/Змішувальна_змінна): *X* та *Y* можуть корелювати не через існування причиново-наслідкового зв'язку між ними, а через те, що обидві залежать від третьої змінної, *Z*. *Z* називають змішувальним фактором ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *confounding factor*).

### Неналежна інтерпретація: кореляція

Поняття [кореляції](https://uk.wikipedia.org/wiki/Кореляція) особливо привертає увагу потенційною плутаниною, яку воно може спричинювати. Статистичний аналіз [набору даних](https://uk.wikipedia.org/wiki/Набір_даних) часто виявляє, що дві змінні (властивості) досліджуваної генеральної сукупності мають схильність змінюватися разом, так, ніби вони пов'язані. Наприклад, дослідження річного доходу, яке також дивиться на вік смерті, може виявити, що тривалість життя в бідних людей схильна бути меншою, ніж у заможних. Про ці дві змінні кажуть, що вони корелюють, проте, вони можуть бути, а можуть і не бути причиною одна одної. Явище кореляції може бути спричинено третім, раніше не розгляданим явищем, що називають неявною змінною ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *lurking variable*) або [змішувальною змінною](https://uk.wikipedia.org/wiki/Змішувальна_змінна) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *confounding variable*). З цієї причини немає можливості негайно зробити висновок про існування причинно-наслідкового зв'язку між цими двома змінними. (Див. [корелювання не означає спричинювання](https://uk.wikipedia.org/wiki/Корелювання_не_означає_спричинювання).)

## Застосування

### Прикладна статистика, теоретична статистика та математична статистика

*Прикладна статистика* включає описову статистику та застосування індуктивної статистики.[[66]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-66)[[67]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-67) *Теоретична статистика* розглядає логічні аргументи, що лежать в основі обґрунтування підходів до [статистичного висновування](https://uk.wikipedia.org/wiki/Статистичне_висновування), а також охоплює *математичну статистику*. Математична статистика включає не лише маніпулювання [розподілами ймовірності](https://uk.wikipedia.org/wiki/Розподіл_імовірності), необхідними, щоби виводити результати, пов'язані з методами оцінювання та висновування, але також і різні аспекти [обчислювальної статистики](https://uk.wikipedia.org/w/index.php?title=Обчислювальна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Computational_statistics) та [планування експериментів](https://uk.wikipedia.org/wiki/Планування_експериментів).

[Статистичні консультанти](https://uk.wikipedia.org/w/index.php?title=Статистичний_консультант&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Statistical_consultant) можуть допомагати організаціям та компаніям, які не мають власної компетенції, що стосується їхніх конкретних питань.

### Машинне навчання та добування даних

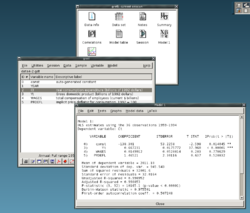
Моделі [машинного навчання](https://uk.wikipedia.org/wiki/Машинне_навчання) — це статистичні та ймовірнісні моделі, що фіксують закономірності в даних через застосування обчислювальних алгоритмів.

### Статистика в академічнім середовищі

Статистику застосовують до широкого спектру [академічних дисциплін](https://uk.wikipedia.org/wiki/Академічна_дисципліна), включно з [природничими](https://uk.wikipedia.org/wiki/Природничі_науки) та [суспільними науками](https://uk.wikipedia.org/wiki/Суспільні_науки), урядуванням та бізнесом. Бізнесова статистика застосовує статистичні методи в [економетрії](https://uk.wikipedia.org/wiki/Економетрія), [аудиті](https://uk.wikipedia.org/wiki/Аудит), виробництві та операційній діяльності, включно з удосконаленням обслуговування та маркетинговими дослідженнями.[[68]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-68) В галузі біологічних наук 12-ма найчастішими статистичними критеріями є: [дисперсійний аналіз](https://uk.wikipedia.org/wiki/Дисперсійний_аналіз) ([англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *ANOVA*), [критерій хі-квадрат](https://uk.wikipedia.org/wiki/Критерій_хі-квадрат), [*t*-критерій Стьюдента](https://uk.wikipedia.org/wiki/T-критерій_Стьюдента), [лінійна регресія](https://uk.wikipedia.org/wiki/Лінійна_регресія), [коефіцієнт кореляції Пірсона](https://uk.wikipedia.org/wiki/Коефіцієнт_кореляції_Пірсона), [*U*-критерій Манна — Уітні](https://uk.wikipedia.org/wiki/U-критерій_Манна—Уітні), [критерій Краскела — Уоліса](https://uk.wikipedia.org/w/index.php?title=Критерій_Краскела_—_Уоліса&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Kruskal-Wallis_Test), [індекс різноманітності Шеннона](https://uk.wikipedia.org/wiki/Індекс_різноманітності_Шеннона), [критерій Тьюкі](https://uk.wikipedia.org/w/index.php?title=Критерій_Тьюкі&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Tukey's_test), [кластерний аналіз](https://uk.wikipedia.org/wiki/Кластерний_аналіз), [коефіцієнт кореляції рангу Спірмена](https://uk.wikipedia.org/wiki/Коефіцієнт_кореляції_рангу_Спірмена) та [метод головних компонент](https://uk.wikipedia.org/wiki/Метод_головних_компонент).[[69]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-:0-69)

Типовий курс статистики охоплює описову статистику, ймовірність, біноміальний та [нормальний розподіли](https://uk.wikipedia.org/wiki/Нормальний_розподіл), перевірку гіпотез та довірчих інтервалів, [лінійну регресію](https://uk.wikipedia.org/wiki/Лінійна_регресія) та кореляцію.[[70]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-70) Сучасні фундаментальні курси статистики для студентів останнього курсу зосереджуються на правильному обиранні критеріїв, інтерпретації результатів, та застосуванні [безкоштовного статистичного програмного забезпечення](https://uk.wikipedia.org/w/index.php?title=Безкоштовне_статистичне_програмне_забезпечення&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Free_statistical_software).[[69]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-:0-69)

### Статистичні обчислення

[](https://uk.wikipedia.org/wiki/Файл:Gretl_screenshot.png)[gretl](https://uk.wikipedia.org/w/index.php?title=Gretl&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/gretl), приклад [відкритого статистичного пакету](https://uk.wikipedia.org/w/index.php?title=Перелік_відкритих_статистичних_програмних_пакетів&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_open_source_statistical_packages)

*Детальніші відомості з цієї теми ви можете знайти в статті* [***Обчислювальна статистика***](https://uk.wikipedia.org/w/index.php?title=Обчислювальна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Computational_statistics)*.*

На практику статистичної науки, починаючи з другої половини XX сторіччя, суттєво вплинуло швидке й стійке зростання обчислювальної потужності. Ранні статистичні моделі майже завжди походили з класу [лінійних моделей](https://uk.wikipedia.org/wiki/Лінійна_модель), але потужні комп'ютери, в поєднанні з відповідними чисельними [алгоритмами](https://uk.wikipedia.org/wiki/Алгоритм), спричинили зростання зацікавлення [нелінійними моделями](https://uk.wikipedia.org/w/index.php?title=Нелінійна_регресія&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Nonlinear_regression) (такими як [нейронні мережі](https://uk.wikipedia.org/wiki/Штучна_нейронна_мережа)), а також створення нових типів, таких як [узагальнені лінійні](https://uk.wikipedia.org/w/index.php?title=Узагальнена_лінійна_модель&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Generalized_linear_model) та [багаторівневі моделі](https://uk.wikipedia.org/w/index.php?title=Багаторівнева_модель&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Multilevel_model).

Збільшення обчислювальної потужності також призвело до зростання популярності обчислювально містких методів, що ґрунтуються на [перевибиранні](https://uk.wikipedia.org/w/index.php?title=Перевибирання_(статистика)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Resampling_(statistics)), таких як [пермутаційний тест](https://uk.wikipedia.org/wiki/Пермутаційний_тест) та [натяжка](https://uk.wikipedia.org/wiki/Статистична_натяжка), тоді як такі методики, як [вибирання за Ґіббсом](https://uk.wikipedia.org/w/index.php?title=Вибирання_за_Ґіббсом&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Gibbs_sampling), зробили здійсненнішим застосування [баєсових моделей](https://uk.wikipedia.org/wiki/Баєсова_модель). Комп'ютерна революція має наслідки для майбутнього статистики з новим акцентом на «експериментальній» та «емпіричній» статистиці. Наразі є доступною велика кількість [статистичного програмного забезпечення](https://uk.wikipedia.org/w/index.php?title=Перелік_статистичних_програмних_пакетів&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_statistical_software) як загального, так і спеціального призначення. До прикладів доступного програмного забезпечення, здатного до складних статистичних обчислень, належать такі програми як [Mathematica](https://uk.wikipedia.org/wiki/Mathematica), [SAS](https://uk.wikipedia.org/w/index.php?title=SAS_(програмне_забезпечення)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/SAS_(software)), [SPSS](https://uk.wikipedia.org/wiki/SPSS) та [R](https://uk.wikipedia.org/wiki/R_(мова_програмування)).

### Застосування статистики в математиці та мистецтві

Традиційно статистика займалася висновуванням із застосуванням напівстандартизованої методології, що була «обов'язковою для вивчення» в більшості наук.[[*джерело?*](https://uk.wikipedia.org/wiki/Вікіпедія:Посилання_на_джерела)] Ця традиція змінилася із застосуванням статистики в не висновувальних контекстах. На те, що колись було нудною темою, обов'язковою для отримання ступеня в багатьох галузях, тепер дивляться з ентузіазмом.[[*на чию думку?*](https://uk.wikipedia.org/wiki/Вікіпедія:Перевірність)] Висміювану спершу деякими математичними пуристами, її тепер вважають важливою методологією в певних областях.

* В [теорії чисел](https://uk.wikipedia.org/wiki/Теорія_чисел) [точкові діаграми](https://uk.wikipedia.org/wiki/Точкова_діаграма) даних, породжуваних функцією розподілу, може бути перетворювано за допомогою звичних інструментів, які використовують у статистиці, щоби розкривати закономірності, що лежать в їх основі, що може потім вести до гіпотез.
* Методи статистики, включено з передбачувальними методами в [прогнозуванні](https://uk.wikipedia.org/wiki/Прогнозування), поєднують з [теорією хаосу](https://uk.wikipedia.org/wiki/Теорія_хаосу) та [фрактальною геометрією](https://uk.wikipedia.org/wiki/Фрактальна_геометрія), щоби створювати, як вважається, дуже красиві відео.[[*джерело?*](https://uk.wikipedia.org/wiki/Вікіпедія:Посилання_на_джерела)]
* [Живопис процесу](https://uk.wikipedia.org/w/index.php?title=Живопис_процесу&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Process_art) [Джексона Поллока](https://uk.wikipedia.org/wiki/Джексон_Поллок) покладався на живописові експерименти, завдяки яким художньо розкривалися розподіли, що лежать в основі природи.[[*джерело?*](https://uk.wikipedia.org/wiki/Вікіпедія:Посилання_на_джерела)] З появою комп'ютерів для формалізування таких ведених розподілами природних процесів було застосовано статистичні методи для створення та аналізу анімованого відеомистецтва.[[*джерело?*](https://uk.wikipedia.org/wiki/Вікіпедія:Посилання_на_джерела)]
* Методи статистики можуть використовувати предикативно в мистецтві [перформансу](https://uk.wikipedia.org/wiki/Перформанс), як-от у фокусі з гральними картами, заснованому на [марковському процесі](https://uk.wikipedia.org/wiki/Марковський_процес), що працює лише деякий час, настання якого можливо передбачувати, використовуючи статистичну методологію.
* Статистику можливо використовувати для предикативного створювання творів мистецтва, як у статистичній або [стохастичній музиці](https://uk.wikipedia.org/wiki/Стохастична_музика), винайденій [Янісом Ксенакісом](https://uk.wikipedia.org/wiki/Яніс_Ксенакіс), де музика є залежною від кожного виконання. І хоч цей тип мистецтва не завжди виходить таким, як очікувалося, він поводиться таким чином, що є передбачуваним та налаштовуваним із застосуванням статистики.

## Спеціалізовані дисципліни

*Докладніше:* [*Список областей застосування статистики*](https://uk.wikipedia.org/wiki/Список_областей_застосування_статистики)

Статистичні методики використовують в широкому спектрі типів наукових та суспільних досліджень, включно з [біологічною статистикою](https://uk.wikipedia.org/wiki/Біологічна_статистика), [обчислювальною біологією](https://uk.wikipedia.org/wiki/Обчислювальна_біологія), [обчислювальною соціологією](https://uk.wikipedia.org/wiki/Обчислювальна_соціологія), [мережною біологією](https://uk.wikipedia.org/w/index.php?title=Мережна_біологія&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Network_biology), [суспільствознавством](https://uk.wikipedia.org/wiki/Суспільствознавство), [соціологією](https://uk.wikipedia.org/wiki/Соціологія) та [суспільствознавчими дослідженнями](https://uk.wikipedia.org/wiki/Суспільствознавчі_дослідження). В деяких областях досліджень прикладну статистику використовують настільки широко, що вони мають [спеціалізовану термінологію](https://uk.wikipedia.org/wiki/Спеціалізована_термінологія). До цих дисциплін належать:

* [Актуарна математика](https://uk.wikipedia.org/wiki/Актуарна_математика) (оцінює ризик в страховій та фінансовій галузях)
* [Астрономічна статистика](https://uk.wikipedia.org/w/index.php?title=Астрономічна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Astrostatistics) (статистичне оцінювання астрономічних даних)
* [Біологічна статистика](https://uk.wikipedia.org/wiki/Біологічна_статистика)
* [Географія](https://uk.wikipedia.org/wiki/Географія) та [геоінформаційні системи](https://uk.wikipedia.org/wiki/Геоінформаційна_система), конкретно в [просторовім аналізі](https://uk.wikipedia.org/wiki/Просторовий_аналіз)
* [Демографія](https://uk.wikipedia.org/wiki/Демографія) (статистичне дослідження населення)
* [Добування даних](https://uk.wikipedia.org/wiki/Добування_даних) (застосування статистики та [розпізнавання образів](https://uk.wikipedia.org/wiki/Розпізнавання_образів) для виявляння знань в даних)
* [Економетрія](https://uk.wikipedia.org/wiki/Економетрія) (статистичний аналіз економічних даних)
* [Епідеміологія](https://uk.wikipedia.org/wiki/Епідеміологія) (статистичний аналіз захворюваності)
* [Медична статистика](https://uk.wikipedia.org/wiki/Медична_статистика)[[71]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-71)
* [Наука про дані](https://uk.wikipedia.org/wiki/Наука_про_дані)
* [Обробка зображень](https://uk.wikipedia.org/wiki/Обробка_зображень)
* [Політологія](https://uk.wikipedia.org/wiki/Політологія)
* Прикладна інформаційна економіка
* [Психологічна статистика](https://uk.wikipedia.org/w/index.php?title=Психологічна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Psychological_statistics)
* [Соціальна статистика](https://uk.wikipedia.org/w/index.php?title=Соціальна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Social_statistics)
* [Статистика енергетики](https://uk.wikipedia.org/w/index.php?title=Статистика_енергетики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Statistical_study_of_energy_data)
* [Статистична механіка](https://uk.wikipedia.org/wiki/Статистична_механіка)
* [Техніка забезпечення надійності](https://uk.wikipedia.org/w/index.php?title=Техніка_забезпечення_надійності&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Reliability_engineering)
* [Хемометрія](https://uk.wikipedia.org/wiki/Хемометрія) (для аналізу даних у [хімії](https://uk.wikipedia.org/wiki/Хімія))
* [Юриметрія](https://uk.wikipedia.org/w/index.php?title=Юриметрія&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Jurimetrics) ([право](https://uk.wikipedia.org/wiki/Право))
* [Фізичне виховання](https://uk.wikipedia.org/wiki/Фізичне_виховання) та [спорт](https://uk.wikipedia.org/wiki/Спорт)[[72]](https://uk.wikipedia.org/wiki/Статистика" \l "cite_note-72)

Крім того, існують певні типи статистичного аналізу, в яких також було розроблено свою власну спеціалізовану термінологію та методологію:

* [Аналіз виживаності](https://uk.wikipedia.org/wiki/Аналіз_виживаності)
* [Багатовимірна статистика](https://uk.wikipedia.org/wiki/Багатовимірна_статистика)
* [Методологія опитування](https://uk.wikipedia.org/wiki/Методологія_опитування)
* [Моделювання структурними рівняннями](https://uk.wikipedia.org/w/index.php?title=Моделювання_структурними_рівняннями&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Structural_equation_modeling)
* [Натяжкова](https://uk.wikipedia.org/wiki/Статистична_натяжка) / [складано-ножева](https://uk.wikipedia.org/wiki/Складано-ножева_перевибірка) [перевибірка](https://uk.wikipedia.org/w/index.php?title=Перевибірка_(статистика)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Resampling_(statistics))
* [Статистична класифікація](https://uk.wikipedia.org/wiki/Статистична_класифікація)
* [Структурний аналіз даних](https://uk.wikipedia.org/w/index.php?title=Структурний_аналіз_даних_(статистика)&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Structured_data_analysis_(statistics))
* Статистика в багатьох видах спорту, зокрема, у [бейсболі](https://uk.wikipedia.org/w/index.php?title=Бейсбольна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Baseball_statistics), відома як [саберметрика](https://uk.wikipedia.org/w/index.php?title=Саберметрика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Sabermetrics), та [крикеті](https://uk.wikipedia.org/w/index.php?title=Крикетна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Cricket_statistics)
* [Методи Тагучі](https://uk.wikipedia.org/wiki/Методи_Тагучі)

Статистика також є ключовим базовим інструментом у бізнесі та виробництві. Її використовують для розуміння варіативності систем вимірювання, керування процесами (як у [статистичнім керуванні процесами](https://uk.wikipedia.org/wiki/Статистичне_керування_процесами), [англ.](https://uk.wikipedia.org/wiki/Англійська_мова) *statistical process control, SPC*), узагальнювання даних, та для здійснення керованих даними рішень. В цих ролях вона є ключовим, і, можливо, єдиним надійним інструментом.

## Див. також

*Докладніше:* [*Нарис теорії статистики*](https://uk.wikipedia.org/wiki/Нарис_теорії_статистики)

* [Оцінювання чисельності](https://uk.wikipedia.org/w/index.php?title=Оцінювання_чисельності&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Abundance_estimation)
* [Наука про дані](https://uk.wikipedia.org/wiki/Наука_про_дані)
* [Глосарій теорії ймовірностей та статистики](https://uk.wikipedia.org/w/index.php?title=Глосарій_теорії_ймовірностей_та_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Glossary_of_probability_and_statistics)
* [Список академічних статистичних асоціацій](https://uk.wikipedia.org/w/index.php?title=Список_академічних_статистичних_асоціацій&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_academic_statistical_associations)
* [Список важливих публікацій в області статистики](https://uk.wikipedia.org/w/index.php?title=Список_важливих_публікацій_в_області_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_important_publications_in_statistics)
* [Перелік національних та міжнародних статистичних служб](https://uk.wikipedia.org/w/index.php?title=Перелік_національних_та_міжнародних_статистичних_служб&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_national_and_international_statistical_services)
* [Список статистичних пакетів](https://uk.wikipedia.org/w/index.php?title=Список_статистичних_пакетів&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_statistical_packages) (програмного забезпечення)
* [Список статей зі статистики](https://uk.wikipedia.org/w/index.php?title=Список_статей_зі_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_statistics_articles)
* [Список університетських статистичних консультаційних центрів](https://uk.wikipedia.org/w/index.php?title=Список_університетських_статистичних_консультаційних_центрів&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_university_statistical_consulting_centers)
* [Позначення в теорії ймовірностей та статистиці](https://uk.wikipedia.org/w/index.php?title=Позначення_в_теорії_ймовірностей_та_статистиці&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Notation_in_probability_and_statistics)
* [Всесвітній день статистики](https://uk.wikipedia.org/wiki/Всесвітній_день_статистики)

Засади та головні області статистики

* [Засади статистики](https://uk.wikipedia.org/w/index.php?title=Засади_статистики&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Foundations_of_statistics)
* [Список статистиків](https://uk.wikipedia.org/w/index.php?title=Список_статистиків&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/List_of_statisticians)
* [Офіційна статистика](https://uk.wikipedia.org/w/index.php?title=Офіційна_статистика&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Official_statistics)
* [Багатовимірний дисперсійний аналіз](https://uk.wikipedia.org/w/index.php?title=Багатовимірний_дисперсійний_аналіз&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/Multivariate_analysis_of_variance)

## Література

* [*Карташов М. В.*](https://uk.wikipedia.org/wiki/Карташов_Микола_Валентинович) [Імовірність, процеси, статистика](http://probability.univ.kiev.ua/userfiles/kmv/VPS_Pv.pdf). — [Київ](https://uk.wikipedia.org/wiki/Київ) : ВПЦ [Київський університет](https://uk.wikipedia.org/wiki/Київський_національний_університет_імені_Тараса_Шевченка), 2007. — 504 с.
* Lydia Denworth, «A Significant Problem: Standard scientific methods are under fire. Will anything change?», [*Scientific American*](https://uk.wikipedia.org/wiki/Scientific_American), vol. 321, no. 4 (October 2019), pp. 62–67. (англ.) «Використання [*p*-значень](https://uk.wikipedia.org/wiki/P-значення) протягом майже сторіччя [з 1925 року] для визначення [статистичної значущості](https://uk.wikipedia.org/wiki/Статистична_значущість) результатів [експериментів](https://uk.wikipedia.org/wiki/Експеримент) посприяло ілюзії [впевненості](https://uk.wikipedia.org/wiki/Упевненість) та [кризі відтворюваності](https://uk.wikipedia.org/wiki/Криза_відтворюваності) в багатьох [наукових галузях](https://uk.wikipedia.org/wiki/Наука). Існує все більша рішучість до реформування статистичного аналізу… Деякі [дослідники] пропонують змінювати статистичні методи, тоді як інші покінчили би з порогом для визначення „значущих“ результатів.» (с. 63)
* Barbara Illowsky; Susan Dean (2014). [*Introductory Statistics*](https://web.archive.org/web/20210721125324/https://openstax.org/details/introductory-statistics). OpenStax CNX. [*ISBN*](https://uk.wikipedia.org/wiki/International_Standard_Book_Number) . Архів [*оригіналу*](https://openstax.org/details/introductory-statistics) за 21 липня 2021. Процитовано 13 вересня 2020. (англ.)
* Stockburger, David W. [*Introductory Statistics: Concepts, Models, and Applications*](https://web.archive.org/web/20200528093101/http://psychstat3.missouristate.edu/Documents/IntroBook3/sbk.htm). [*Missouri State University*](https://uk.wikipedia.org/w/index.php?title=Університет_штату_Міссурі&action=edit&redlink=1)[*[en]*](https://en.wikipedia.org/wiki/Missouri_State_University) (вид. 3rd Web). Архів [*оригіналу*](http://psychstat3.missouristate.edu/Documents/IntroBook3/sbk.htm) за 28 травня 2020. (англ.)
* [*OpenIntro Statistics*](https://www.openintro.org/stat/textbook.php?stat_book=os) [[Архівовано](https://web.archive.org/web/20190616110442/https://www.openintro.org/stat/textbook.php?stat_book=os) 16 червня 2019 у [Wayback Machine](https://uk.wikipedia.org/wiki/Wayback_Machine).], 3rd edition by Diez, Barr, and Cetinkaya-Rundel (англ.)
* Stephen Jones, 2010. [*Statistics in Psychology: Explanations without Equations*](https://books.google.com/books?id=mywdBQAAQBAJ) [[Архівовано](https://web.archive.org/web/20200727151017/https://books.google.com/books?id=mywdBQAAQBAJ) 27 липня 2020 у [Wayback Machine](https://uk.wikipedia.org/wiki/Wayback_Machine).]. Palgrave Macmillan. [ISBN 9781137282392](https://uk.wikipedia.org/wiki/Спеціальна:Джерела_книг/9781137282392). (англ.)
* Cohen, J (1990). [*Things I have learned (so far)*](https://web.archive.org/web/20171018181831/http://moityca.com.br/pdfs/Cohen_1990.pdf) (PDF). American Psychologist. **45**: 1304—1312. [*doi*](https://uk.wikipedia.org/wiki/Цифровий_ідентифікатор_об'єкта):[*10.1037/0003-066x.45.12.1304*](https://doi.org/10.1037%2F0003-066x.45.12.1304). Архів [*оригіналу*](http://moityca.com.br/pdfs/Cohen_1990.pdf) (PDF) за 18 жовтня 2017. (англ.)
* Gigerenzer, G (2004). Mindless statistics. Journal of Socio-Economics. **33**: 587—606. [*doi*](https://uk.wikipedia.org/wiki/Цифровий_ідентифікатор_об'єкта):[*10.1016/j.socec.2004.09.033*](https://doi.org/10.1016%2Fj.socec.2004.09.033). (англ.)
* Ioannidis, J.P.A. (2005). [*Why most published research findings are false*](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1855693). PLoS Medicine. **2**: 696—701. [*doi*](https://uk.wikipedia.org/wiki/Цифровий_ідентифікатор_об'єкта):[*10.1371/journal.pmed.0040168*](https://doi.org/10.1371%2Fjournal.pmed.0040168). [*PMC*](https://uk.wikipedia.org/wiki/PubMed_Central) [*1855693*](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1855693). [*PMID*](https://uk.wikipedia.org/wiki/PMID) [*17456002*](https://pubmed.ncbi.nlm.nih.gov/17456002). (англ.)
* Нариси з історії статистики України / Авт. кол.: О. Г. Осауленко (гол. ред.), В. І. Карпов, М. В. Пугачова та ін.; Держкомстат України, НДІ статистики. — 2-е вид., випр. та доп. — К., 1999. — 188 с.: іл. — Бібліогр.: с. 147—152 (133 назви).
* Статистика: навчальний посібник / С. О. Матковский, М. Л. Вдовин, Т. В. Панчишин. — Львів: Видавничий центр ЛНУ імені Івана Франка, 2010. — 344 с.
* Статистика: навч. посіб. для студ. екон. спец. вищ. навч. закл. / С. О. Матковський, Л. І. Гальків, О. С. Гринькевич, О. З. Сорочак. — 2-ге вид., доповн. і виправл. — Л. : Новий Світ-2000, 2011. — 429, [3] с. : іл. — (Вища освіта в Україні). — Бібліогр. в кінці розділів. — [ISBN 978-966-418-089-1](https://uk.wikipedia.org/wiki/Спеціальна:Джерела_книг/9789664180891)
* Статистика: навч. посіб. / Р. В. Фещур, В. П. Кічор, А. Ф. Барвінський, М. Р. Тимощук ; М-во освіти і науки України, Нац. ун-т «Львів. політехніка». — 4-те вид., оновл. і доповн. — Л. : Бух. центр «Ажур», 2010. — 256 с. : іл., табл. — Бібліогр.: с. 251—252 (20 назв). — [ISBN 978-966-1688-05-5](https://uk.wikipedia.org/wiki/Спеціальна:Джерела_книг/9789661688055)
* Статистика ринку товарів та послуг: Навч. посіб. / Л. І. Крамченко; Укоопспілка. — Л., 2002. — 188 c. — Бібліогр.: 32 назви.