

Практична робота № 3 Вибірковий метод. Точкові оцінки вибірових статистичних показників

Мета роботи: отримання практичних навичок з визначення точкових оцінок статистичних показників вибіркової сукупності.

Програма виконання роботи

1. Завантажити табличний процесор MS Excel.
2. Вибрати із табл. результати вимірювань трьох довільних показників якості ґрунту і оформити їх у вигляді таблиці на аркуші Excel.
3. Під останнім значенням першого ряду спостережень за допомогою вбудованих функцій категорії **Статистические: СЧЕТ, СРЗНАЧ, ДИСПР, СТАНДОТКЛОНП, МАКС, МИН, СКОС, ЭКСЦЕСС** визначити відповідні статистичні показники, розташовуючи їх один під одним.
4. Визначити вибірковий коефіцієнт варіації за формулою (5).
5. Розповсюдити формули на дві комірки праворуч, щоб визначити відповідні показники для інших вибраних показників ґрунту.
6. Оцінити відповідність експериментальних розподілів нормальному за допомогою значень асиметрії та ексцесу (чи знаходяться відповідні значення в межах $\pm 0,5$).
7. Скопіювати одержані результати на новий аркуш.

8. На новому аркуші визначити показники аномальності для мінімального і максимального значень вибірки за формулами (11.8).
9. Вилучити аномальні значення з вибірки, якщо такі виявлено.
10. Повторити визначення показників аномальності для наступних мінімального і/або максимального значень вибірки.
11. Вилучити аномальні значення з вибірки, якщо такі виявлені повторно.
12. Повторювати перевірку даних на аномальність і їх вилучення до виконання умови $V_{\max, \min} \leq 3$.
13. Після вилучення всіх виявлених аномальних експериментальних значень звернути увагу на те, як зміняться статистичні числові показники вибірки. Повторно оцінити відповідність експериментального розподілу нормальному розподілу за допомогою значень асиметрії та ексцесу.
14. Порівняти основні статистичні числові характеристики отримані до і після вилучення аномальних даних.
15. Відмітити, які показники при цьому змінилися.
16. Дати пояснення отриманим результатам.
17. Зберегти документ в папці під своїм прізвищем, яку помістити в папку **"Мои документы"**.

Застосування вбудованих функцій MS Excel для визначення статистичних показників вибіркової сукупності

У MS Excel основні статистичні характеристики вибіркової сукупності визначаються за допомогою вбудованих функцій категорії **Статистические**. Для їх використання виконують такі дії:

- **Вставка=>Функция**;
- або користуються піктограмою fx (**Вставка функции**), яка розташована на панелі інструментів;
- у діалоговому вікні **Мастер функций**, що відкриється, у полі **Категория** вибирають **Статистические**;
- у полі **Выберите функцию** вибирають потрібну функцію;
- у діалоговому вікні **Аргументы функции**, що відкриється після вибору функції, вводять діапазон комірок з вхідними даними, або перевіряють, чи правильно він заданий, якщо діапазон комірок був попередньо вибраний;
- натискають кнопку **ОК**.

Перелік і призначення статистичних функцій наведено в табл.

Вид функції	Призначення
СЧЕТ	Обсяг вибірки
СРЗНАЧ	Вибіркове середнє значення
ДИСПР	Вибіркова дисперсія
СТАНДОТКЛОНП	Вибіркове середнє квадратичне відхилення
МАКС	Максимальне число вибіркової сукупності
МИН	Мінімальне число вибіркової сукупності
СКОС	Показник асиметрії A
ЭКСЦЕСС	Показник ексцесу E

Середнє вибіркве значення $\bar{x}_b$ (середнє арифметичне значення ознаки вибіркової сукупності, що досліджується) визначається за формулою

$$\bar{x}_b = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

x_i – значення i -ого вимірювання, n – кількість вимірювань.

Вибіркова дисперсія D_b (середнє арифметичне квадратів відхилень значень ознаки x_i що досліджується, від середнього вибіркового значення $\bar{x}_b$) визначається за формулою

$$D_b = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_b)^2 \quad (11.3)$$

Вибіркове середнє квадратичне відхилення σ_b визначається за формулою

$$\sigma_b = \sqrt{D_b} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_b)^2} \quad (11.4)$$

Для оцінки варіації даних використовують вибірковий коефіцієнт варіації C_b , який обчислюється за формулою

$$C_b = \frac{\sigma_b}{\bar{x}_b} \cdot 100\% \quad (11.5)$$

Коефіцієнт варіації C_b застосовують для порівняння варіації рядів спостережень, що відрізняються середніми значеннями і дисперсією. Окрім цього, C_b - величина безрозмірна і може використовуватися для порівняння варіації рядів спостережень, що мають різні одиниці вимірювання.

Оцінка відхилення експериментального розподілу від нормального

Обчислення статистичних показників правомірне за умови підпорядкування експериментальних даних нормальному закону розподілу. При вивченні невідомих (експериментальних) розподілів, або розподілів, що відрізняються від нормального, виникає потреба кількісно оцінити цю відмінність. З цією метою застосовують спеціальні характеристики, зокрема асиметрію (коефіцієнт асиметрії) і ексцес (коефіцієнт ексцесу). Асиметрія показує, наскільки розподіл даних несиметричний відносно нормального розподілу. Якщо асиметрія є величиною додатною, то більша частина даних має значення, що перевищує середнє вибіркове $\bar{x}_v$. Якщо асиметрія менше нуля, то більша частина даних має значення менше за $\bar{x}_v$. Ексцес оцінює крутість, тобто величину більшого, або меншого підйому вершини графіка розподілу експериментальних даних порівняно з вершиною графіка нормального розподілу. Якщо ексцес є величиною додатною, то вершина графіка експериментального розподілу вище нормального, якщо ексцес менше нуля, то - нижче нормального. Для нормального розподілу ці показники дорівнюють нулю.

Якщо для розподілу, що вивчається, асиметрія і ексцес знаходяться в межах $\pm 0,5$, то можна припустити, що експериментальний закон розподілу близький до нормального. При цьому припускають, що емпіричний і теоретичний нормальні розподіли мають однакові математичне сподівання (середнє) і дисперсію. За умови малої кількості спостережень, перед використанням асиметрії та ексцесу для оцінки близькості експериментального розподілу до нормального треба оцінити точність визначення вказаних характеристик.

Асиметрія A визначається за формулою

$$A = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_B)^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_B)^2 \right)^{\frac{3}{2}}} \quad (11.6)$$

Ексцес E визначається за формулою

$$E = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_B)^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_B)^2 \right)^2} - 3 \quad (11.7)$$

Визначення вибірових статистичних показників множини ознак, що досліджуються

Якщо об'єкт, процес, або система, що досліджуються, характеризуються не однією ознакою, а множиною ознак, які представлені своїми рядами спостережень, то для визначення множини статистичних показників виконують дії:

- визначають статистичний показник однієї ознаки за допомогою відповідної вбудованої функції категорії **Статистические**;
- розповсюджують, або копіюють вбудовану функцію на діапазон комірок, що потребують аналогічного виду обчислень.

Виявлення і вилучення аномальних даних

Однією із причин, коли за показниками асиметрії і ексцесу експериментальний закон розподілу не можна вважати близьким до нормального, є наявність у вибірковій сукупності аномальних даних, тобто таких даних, які за своїми значеннями різко відрізняються від решти. Такі дані не можна віднести до нормально розподілених. У разі підозри на наявність аномальних даних перевірці підлягають максимальне і мінімальне вибіркові значення.

Показники аномальності $V_{\max}$ і $V_{\min}$ визначають за формулами

$$V_{\max} = \frac{(x_{\max} - \bar{x}_e)}{\sigma_e}, \quad V_{\min} = \frac{(\bar{x}_e - x_{\min})}{\sigma_e}, \quad (11.8)$$

де $x_{\max}$, $x_{\min}$ – відповідно максимальне і мінімальне значення експериментальних даних; $\bar{x}_v$ - вибіркове середнє значення; σ_v – вибіркове середнє квадратичне відхилення.

Для перевірки аномальності даних застосовують правило трьох сигм: якщо випадкова величина розподілена нормально, то абсолютна величина її відхилення від математичного сподівання не перевищує потроєного середнього квадратичного відхилення. Тобто, якщо показники аномальності $V_{\max, \min} \leq 3$, то відповідне максимальне і/або мінімальне значення не є аномальним і з подальшого аналізу не вилучається. Якщо $V_{\max, \min} > 3$, то відповідне максимальне і/або мінімальне значення є аномальним і його з подальших досліджень виключають.

Після вилучення аномальних даних числові статистичні показники вибірки обчислюють повторно і перевіряють наступні максимальне і мінімальне значення на аномальність. Якщо при розрахунках за формулами використовувати посилання на адреси комірок з потрібними значеннями, то в Excel перерахунок показників виконається автоматично.