

1.2. ПЕРЕВІРКА ОДНОРІДНОСТІ ВИБІРКИ. ВИЯВЛЕННЯ ЗНАЧЕНЬ ВАРІАНТ, ЩО ВИПАДАЮТЬ

Як було зазначено вище, значення $\bar{X}$, s^2 , s і s_x можуть бути визнані вірогідними, якщо жодна з варіант вибірки не обтяжена грубою похибкою, тобто якщо вибірка однорідна.

Причини появи у вибірці грубих похибок (outliers) можуть бути різні: недостатня кваліфікація персоналу, збої в роботі обладнання, нетипова неоднорідність зразка тощо. Після виявлення грубих похибок, ці результати далі не використовуються під час розрахунку статистичних характеристик. Але про їх наявність має бути повідомлено, і причини їх появи мають бути з'ясовані.

Виявлення грубих похибок (outliers) — дуже складне завдання, щодо якого в літературі немає єдиної усталеної думки (див. розділ 7.3 статті 5.3 «Статистичний аналіз результатів біологічних випробувань і кількісних визначень»). Це особливо стосується вибірок, зовсім малих за обсягом (3-5 вимірювань).

У разі виявлення грубих похибок зазвичай передбачається нормальний розподіл варіант навколо середнього значення (Гаусс або його узагальнений варіант — Стьюдент). Згідно з цим розподілом, будь-яке велике відхилення окремої варіанти (зокрема, і грубої похибки) від середнього значення має ймовірність, яка відрізняється від нуля. Тому для виявлення грубих похибок ми повинні визначити якусь прийнятну для нас ймовірність (рівень значущості α) того, що варіанта належить до генеральної сукупності. Якщо ймовірність нижче ~~нього~~ рівня значущості α , то вважається, що ця варіанта є грубою похибкою. Зазвичай рівень $\alpha = 0.05$ вважають значущим, а $\alpha = 0.01$ — дуже значущим для того, щоб визнати варіанту грубою похибкою.

З/і ростом/зростанням обсягу вибірки (n) змінюються підходи до виявлення грубих похибок. Тому (за обсягом) вибірки можна умовно поділити на:

- *Малі за обсягом ($n \leq 8$).* Ці вибірки є типовими під час проведення рутинного контролю якості (особливо $n = 3 \dots 5$). Зокрема, саме в таких межах варіюється кількість паралельних інжекцій під час проведення випробування на придатність хроматографічної системи (див. Табл. 2.2.46-2 розділу «Методи хроматографічного розділення» (2.2.46)). Для цих вибірок наявність грубих похибок суттєво спотворює статистичні характеристики. Тому перевірку однорідності малих вибірок і виявлення грубих похибок доцільно здійснювати без попереднього обчислення статистичних характеристик за критерієм Діксона (який зветься також Q -критерієм).
- *Середні за обсягом ($9 \leq n \leq 25$).* Ці вибірки зазвичай зустрічаються під час валідаційних досліджень, внутрішньо- і міжлабораторних тестувань тощо. Перевірку однорідності цих вибірок здійснюють після

попереднього обчислення статистичних характеристик $\bar{x}$ і s , які в цьому разі є значно надійнішими, ніж для малих вибірок. Для виявлення грубих помилок вибірок, середніх за обсягом, зазвичай використовують випробування Граббса.

- *Великі за обсягом ($n > 25$)*. Такі вибірки зазвичай зустрічаються під час валідаційних досліджень технології і міжлабораторних тестувань. Для них суттєвою стає ймовірність великих відхилень від середнього значення. Тому для вибірок такого обсягу доцільно використовувати $3s$ -критерій.

1.2.1. Вибірки, малі за обсягом: критерій Діксона (Q -критерій)

Зазвичай використовують для малих вибірок ($n \leq 8$), хоча дає коректні результати і для $n \leq 10$. Q -критерій ґрунтується на використанні розмаху варіювання R .

Попередньо вибірку представляють у ранжованому вигляді (1.1) і для крайніх варіант x_1 і x_n (які припускаються такими, що випадають) розраховують значення контрольного критерію Q на підставі значення розмаху варіювання R :

$$R = \begin{cases} |x_1 - x_n| & \text{для } n=3 \dots 7 \\ |x_1 - x_{n-1}| & \text{для } n=8 \dots 10 \end{cases} \quad (1.10)$$

$$Q_1 = \frac{|x_1 - x_2|}{R} \quad (1.11a)$$

$$Q_n = \frac{|x_n - x_{n-1}|}{R} \quad (1.11b)$$

Вибірка визнається неоднорідною, якщо хоча б одне з обчислених значень Q_1 чи Q_n перевищує табличне значення $Q(n, \alpha)$, знайдене для рівня значущості α (див. Табл. 10.1.1 Додатка). Варіанти x_1 або x_n , для яких відповідне значення $Q > Q(n, \alpha)$, відкидаються, і для одержаної вибірки зменшеного обсягу виконують новий цикл обчислень за рівняннями 1.10 і 1.11, щоб перевірити її однорідність.

Одержану в кінцевому підсумку однорідну вибірку використовують для обчислення $\bar{x}$, s^2 , s і $s_{\bar{x}}$.

Приклад обчислень наведений у розділі 8.2.

Застосування підходу, що базується на співвідношеннях (1.10-1.11), накладає

обмеження на достатню чутливість шкали вимірювань (див. розділ 1.2.4).

1.2.2. Вибірки, середні за обсягом: випробування Граббса

Випробування Граббса є окремим випадком (виявлення однієї грубої похибки за один раз) більш загального випробування (виявлення декількох грубих похибок за один раз), яке ґрунтується на оцінці екстремальних відхилень, нормалізованих за критерієм Стюдента (Extreme Studentized Deviate Test), — *ESD*-випробування.

Передумовою застосування випробування Граббса є перевірка нормальності розподілу значень варіант. Якщо досліджувана сукупність має інший, особливо асиметричний, розподіл (наприклад, логнормальний), це випробування може давати некоректні висновки.

Випробування Граббса рекомендується використовувати для $9 \leq n \leq 25$, хоча табличні дані охоплюють більший діапазон. За $n > 25$ це випробування стає дуже приблизним.

Як і для *Q*-критерію, вибірку попередньо представляють у ранжованому вигляді (1.1) і розраховують статистичні характеристики $\bar{x}$ і s .

Випробування ґрунтується на оцінці різниці (t) між середнім значенням вибірки $\bar{x}$ і найбільшою (x_n) або найменшою (x_l) варіантою в термінах стандартного відхилення вибірки s .

$$t_{\max} = \frac{x_n - \bar{x}}{s}, \quad t_{\min} = \frac{\bar{x} - x_l}{s} \quad (1.12)$$
$$t_{\max} \text{ або } t_{\min} \leq ESD_{crit}.$$

Тут x_n або x_l — значення крайніх варіант вибірки;

$\bar{x}$ — середнє значення;

s — стандартне відхилення, розраховане для всієї вибірки.

ESD_{crit} — критичне значення критерію Граббса для вибраного рівня значущості (див. Табл. 11.1.2).

Випробування застосовують для виявлення однієї грубої похибки для заданого рівня значущості α у вибірці, для якої припускають нормальний розподіл. Випробування повторюють доти, поки грубі похибки не будуть виявлятися. Водночас мінімальний розмір остаточної вибірки не має бути менше 7, тобто $n \geq 7$.

Приклад обчислень наведений у розділі 8.2.

1.2.3. Вибірки, великі за обсягом: *3s*-критерій

Для вибірок, великих за обсягом ($n > 25$), перевірку однорідності проводять також після попереднього обчислення статистичних характеристик $\bar{x}$ і s . Водночас вибірка визнається однорідною, якщо для всіх варіант відхилення d_i (1.3) задовольняє вимогам $3s$ -критерію:

$$|d_i| \leq 3 \times s. \quad (1.13)$$

Зазначимо, що критерій Граббса (див. Табл. 10.1.2) для $n = 25$ і рівня значущості $\alpha = 0.01$ ($ESD_{crit} = 3.009$) співпадає з критерієм $3s$ і відповідає за однобічним розподілом Стюдента ймовірності $P_l > 99.7\%$.

Якщо вибірка визнана неоднорідною, варіанти, для яких $|d_i| > 3 \times s$, відкидаються як обтяжені грубими похибками. У цьому разі для одержаної вибірки скороченого обсягу повторюють цикл обчислень статистичних характеристик за формулами 1.2-1.7 і знову проводять перевірку однорідності. Обчислення статистичних характеристик вважають закінченим, коли вибірка скороченого обсягу виявляється однорідною.

Приклад застосування $3s$ -критерію (1.13) для виявлення варіант, що випадають, наведений у розділі 8.9.2.

1.2.4. Загальні зауваження щодо виявлення варіант, що випадають

1.2.4.1. Вибірки, малі за обсягом

У разі застосування Q -критерію для малих вибірок існує практична проблема недостатньої чутливості шкали вимірювань. Якщо завдяки занадто грубій шкалі всі варіанти, за винятком одного значення, співпадають, то, незалежно від того, як така варіанта відрізняється від інших, вона буде помилково ідентифікуватися як така, що випадає. Така ситуація може виникати у разі дуже малих вибірок ($n = 3 \dots 5$).

Наприклад, в експерименті були одержані такі значення оптичної густини: 0.4335, 0.4334, 0.4335. Згідно з підходом 1.2.1, величина 0.4334 випадає, що є абсурдним.

Для коректного застосування підходу розділу 1.2.1 необхідно, щоб мінімальний крок шкали вимірювань (*step*) був незначущий порівнюючи з розмахом варіювання R . Тобто, згідно зі співвідношенням (2.6), має виконуватися нерівність:

$$step \leq 0.32 \times R.$$

У разі малих вибірок великою є невизначеність статистичних характеристик, що призводить до високої статистичної незначущості грубих похибок. Так, із Табл. 10.1.1 Додатка видно, що $Q = 0.94$ для $n = 3$ (така кількість паралельних випробувань є звичайною для рутинного аналізу) і $P = 95\%$. Це означає, що якщо в нас, наприклад, значення x_3 значно відрізняється від x_1 і x_2 , то ми

можемо його виключити тільки в тому разі, коли різниця $|x_2 - x_3|$ буде у 20 разів більше, ніж $|x_1 - x_2|$ (див. рівняння (1.10-1.11). У разі $p = 0.99$ така різниця має бути у 100 разів більшою. Ці приклади ілюструють обмеженість на практиці застосування для малих вибірок чисто статистичних критеріїв для ідентифікації варіант, що випадають.

1.2.4.2. Вибірки, великі за обсягом

По мірі збільшення вибірки зменшується невизначеність статистичних характеристик, але й збільшується вірогідність великих відхилень від середнього результату, обумовлених нормальним розподілом варіант. Якщо, наприклад, вірогідність для однієї варіанти потрапити в якийсь довірчий інтервал становить $p_1 = 0.99$, то ймовірність того, що всі n варіант потраплять у цей самий інтервал, дорівнює $p_n = 0.99^n$. Відповідно, вірогідність того, що хоча б одна з цих n варіант не потрапить у цей довірчий інтервал, дорівнює $p_{0,n} = 1 - 0.99^n$. Розрахунки показують, що навіть для вибірки $n = 10$ вірогідність $p_{0,n} = 0.10$ (тобто перевищує прийнятний у статистиці рівень значущості $\alpha = 0.05$), для $n = 25$ маємо $p_{0,n} = 0.22$, а для $n = 80$ маємо $p_{0,n} = 0.55$, тобто більше половини.

Для компенсації цього фактора критичні значення випробування Граббса для вибраного рівня значущості (наприклад, $\alpha = 0.01$) зі збільшенням обсягу вибірки постійно зростають: $ESD_{crit} = 2.410$ ($n = 10$), $ESD_{crit} = 3.009$ ($n = 25$), $ESD_{crit} = 3.411$ ($n = 60$) (див. Табл. 10.1.2). Відповідно, зменшується і дозволена цим критерієм ймовірність для однієї варіанти не потрапити в цей довірчий інтервал: 0.0196 ($n = 10$), 0.0030 ($n = 25$), 0.0006 ($n = 60$).

Усі статистичні розподіли (зокрема, Гаусса і Стюдента) є приблизними, а за $n > 25$ випробування Граббса є дуже приблизним. Особливо це стосується крайніх частин цих розподілів. Тому висновок про віднесення варіанти до грубої похибки (outlier), який ґрунтується на таких малих імовірностях, викликає певні сумніви.

Зважаючи на це, для вибірок $n > 25$ доцільним є застосування $3s$ -критерію, який є умовним: він постулює, що варіанта, яка не відповідає критерію (1.13), є такою, що обтяжена грубою похибкою. Для вибірки $n > 25$ це відповідає ймовірності $p_{0,n} < 0.0030$ того, що ця окрема варіанта належить до генеральної сукупності. У разі $n = 25$ $3s$ -критерій співпадає з критерієм Граббса для рівня значущості 0.01.

Проблема виявлення грубих похибок для великих вибірок може вирішуватися по-різному з огляду на формулювання аналітичного завдання (див. розділ 2.2).

Тому для виявлення варіант, що випадають, за думкою дослідника (наприклад, з погляду загальної аналітичної практики), залучають, крім чисто статистичних, також інші підходи (див. розділ 6.4). Обов'язковою

передумовою використання таких підходів є те, що всі критерії мають бути сформульовані до проведення експерименту.