

ПЕРЕВІРКА КРИТИЧНОСТІ ДАНИХ МІЖЛАБОРАТОРНИХ ВИПРОБУВАНЬ ІЗ ЗАСТОСУВАННЯМ СТАТИСТИЧНИХ МЕТОДІВ

У міжнародних стандартах стандартизовано різноманітні статистичні методи, зокрема графічні, які застосовують для проведення перевірки критичності даних міжлабораторних випробувань. Проведено порівняльний аналіз стандартизованих графічних методів і критеріїв для виявлення викидів у отриманих результатах випробувань. Розглянуто особливості застосування статистичних методів і випробувань на викиди у результатах випробувань.

Ключові слова: статистичний метод, перевірка критичності даних, міжлабораторні випробування, стандарт.

Вступ

При плануванні, проведенні та статистичному аналізі міжлабораторних випробувань для оцінки відмінності при використанні стандартної методики вимірювань та визначенні повторюваності та відтворюваності даних, отриманих під час міжлабораторних випробувань, використовуються стандартизовані статистичні методи [1–3].

При виконанні статистичного аналізу отриманих даних необхідно їх дослідити і застосовувати будь-які статистичні методи для ідентифікації потенційних викидів. З цією метою застосовуються різноманітні стандартизовані статистичні методи, до вибору яких необхідно відноситися досить виважено. Актуальним є проведення порівняльного аналізу можливостей статистичних методів і вироблення необхідних рекомендацій щодо їх застосування.

1. Стандартизовані статистичні методи перевірки критичності первинних даних

Основною метою міжнародного стандарту ISO/TR 22971 [3] є забезпечення його користувачів практичними настановами щодо використання стандарту ISO 5725-2 [4] щодо основного методу визначення повторюваності та відтворюваності стандартного методу вимірювання. У стандарті подано інформацію у вигляді простої послідовної процедури для планування, виконання та статистичного аналізу міжлабораторних випробувань, проведених для оцінювання нестійкості стандартного методу вимірювань і визначення повторності та відтворності даних, одержаних при проведенні

міжлабораторних випробувань.

Стандарт ISO 13528 [5] містить детальні описи статистичних методів, які використовують під час аналізу даних, одержаних за допомогою схем перевірки кваліфікації, і дає рекомендації щодо їхнього використання учасниками таких схем та уповноваженими особами. Його призначають як доповнення до стандарту ISO/IEC 17043 [6] щодо перевірки вмінь за допомогою міжлабораторних випробувань.

Перед проведенням будь-яких випробувань з метою визначення потенційних викидів, рекомендується отримувати графічні діаграми з первинними даними. Розгляд таких діаграм дозволяє виявляти викиди, отримані за результатами міжлабораторних випробувань.

На практиці використовуються різноманітні графічні діаграми, тому доцільно провести їх порівняльний аналіз з метою виявлення особливостей кожної з них.

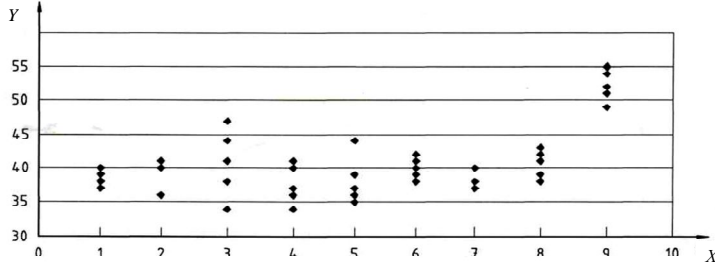
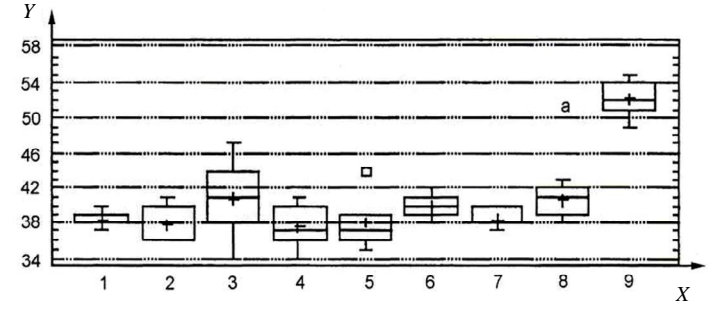
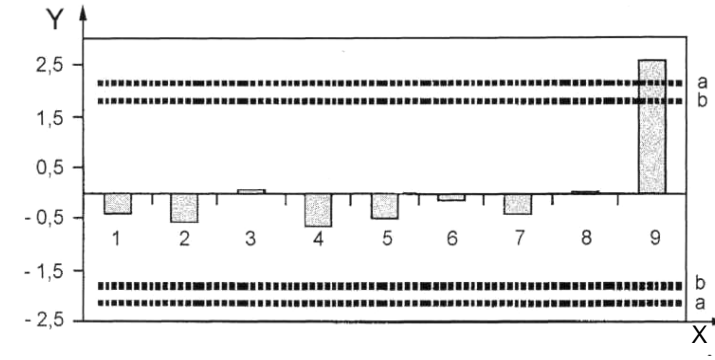
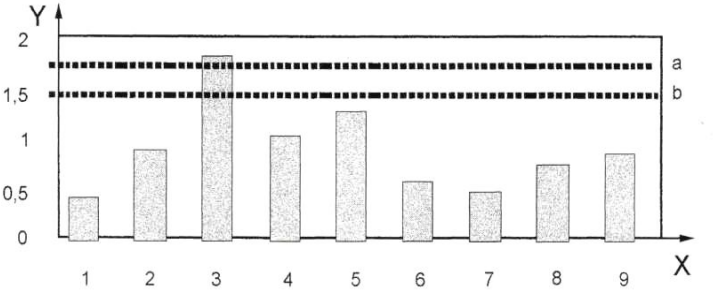
В табл. 1 зведені найбільш розповсюджені діаграми для перевірки критичності первинних даних.

З діаграми виду миттева “картинка” може бути отримана велика кількість інформації, зокрема щодо наявності викидів або очевидної незрозумілої різниці в даних, шляхом візуальної перевірки первинних даних, а також досягається миттева оцінка цих даних (рис. 1).

Діаграма виду “ящик з вусами” (рис. 2) базується на твердій статистиці, включаючи визначення середньої величини.

Значення коефіцієнтів (співвідношень) для h - чи k -статистик Манделя зводяться у відповідну діаграму для всіх лабораторій і для всіх рівнів (рис. 3 чи рис. 4) і порівнюються з розрахованими або табличними коефіцієнтами (співвідношеннями), отриманими для 95 % і 99 % довірчої ймовірності.

Таблиця 1 – Найбільш поширені графічні діаграми для перевірки критичності даних

Назва і загальна характеристика графічної діаграми	Вигляд графічної діаграми
<p>Діаграма виду миттева “картинка” (X – ідентифікація лабораторії; Y – результати в лабораторії), миттева оцінка шляхом простої візуальної перевірки відповідної діаграми отриманих первинних даних</p>	 <p>Рис. 1 Діаграма виду миттева “картинка”</p>
<p>Діаграма виду “ящик з вусами” (“+” – середнє значення; X – ідентифікація лабораторії; Y – результати в лабораторії), отримані середні величини для всіх лабораторій використовують для розрахунку одиничного загального середнього значення</p>	 <p>Рисунок 2 – Діаграма виду “ящик з вусами”</p>
<p>Діаграма h-статистики Манделя (X – ідентифікація лабораторії; Y – результати в лабораторії; a – 99 % довірча ймовірність; b – 95 % довірча ймовірність), співвідношення різниці між середнім значенням для базового елемента та загальним середнім значенням для даного рівня, а також стандартного відхилення середніх значень в базових елементах</p>	 <p>Рисунок 3 – Діаграма h-статистики Манделя</p>
<p>Діаграма k-статистики Манделя (X – ідентифікація лабораторії; Y – результати в лабораторії; a – 99 % довірча ймовірність; b – 95 % довірча ймовірність), співвідношенням стандартного відхилення результатів і середнє або сумарнє стандартнє відхилення</p>	 <p>Рисунок 4 – Діаграма k-статистики Манделя</p>

Аналіз результатів лабораторій, що брали участь у міжлабораторних випробуваннях, наведених на рис. 1–4, показує наступне.

Результати для лабораторії 9 на діаграмах, наведених на рис. 1–3, могли б визначати викиди з врахуванням середнього значення лабораторії при порівнянні з іншим середнім значенням для всіх інших лабораторій. У цьому випадку необхідно оцінити відтворюваність результатів для цієї лабораторії із застосуванням випробувань за критерієм Грабса [3].

Якщо випробування за критерієм Грабса показує наявність викидів, то діаграма для кожної лабораторії для k -статистики може досягати або перевищувати розраховані значення для h -статистики Манделя при 95 % або 99 % довірчої ймовірності.

Результати для лабораторії 3 на діаграмах, наведених на рис. 1–2 і рис. 4, могли б визначати викиди і можна б було забезпечити більш ефективно отримання результатів в порівнянні з іншими лабораторіями. В цьому випадку необхідно оцінити повторюваність результатів для цієї лабораторії із застосуванням випробувань за критерієм Кохрена [3].

Якщо випробування за критерієм Кохрена показує наявність викидів, то діаграма для кожної лабораторії може досягати або перевищувати розраховану статистику Манделя за довірчої ймовірності 95 % або 99 %.

Як викид розглядається кожний результат, який досить сильно відрізняється від усіх інших результатів, що використовуються для подальших досліджень. Залежно від типу розподілу, якому належать результати спостережень, результат, який здається викидом, насправді може ним не бути. Стандартом ISO 5725-2 [4] рекомендується довіряча ймовірність 95 % для викидів, які називають “квазिवикидами”, і 99 % для викидів, які називають “статистичними викидами”.

Вибір 95 % і 99 % довірчої ймовірності для конкретних обставин означає, що один результат із 20 і один результат із 100, відповідно, можуть бути помилково тлумачено. Виходячи з цього, цей єдиний результат може випасти випадково, а встановлений рівень довірчої ймовірності може не відповідати конкретним вимогам. Він може представляти міру прийнятності, якої недостатньо для певних цілей.

Контроль викидів за графіками допомагає отримати додаткове обґрунтування за результатами випробувань на викиди (кількості викидів). Отже, графіки h - і k -статистик Манделя допускається використовувати для полегшення цих рішень. Розгляд повинен дати відхилення

всіх результатів лабораторії, якщо конкретний набір даних (для лабораторії 9 на рис. 3 або для лабораторії 3 на рис. 4) показує, що всі розрахункові значення позитивні та близькі або перевищують наведені в таблицях значення для 95 % і 99 % рівня довірчої ймовірності.

Можливо, що всі вибірккові середні однієї лабораторії більші, ніж відповідні значення всіх інших лабораторій. Цей факт може бути причиною для занепокоєння. Як раніше, рішення про відхилення даних слід приймати лише на основі випробувань за певними критеріями для виявлення вірогідної причини викидів.

2. Особливості застосування критеріїв, які використовуються для перевірки критичності первинних даних

У випробуваннях за певними критеріями, які використовують для визначення наявності або відсутності викидів, передбачають, що результати розподілені відповідно до розподілу Гауса (нормальний розподіл) або, принаймні, відповідно до одномодального розподілу. Отже, перед виконанням будь-якого випробування, особливо випробування, яке вимагає використання великої кількості результатів, має бути зроблена перевірка цього припущення, зокрема з використанням стандарту ISO 5479 [7].

Також передбачається, що кількість результатів в кожному наборі даних (від кожної лабораторії) однакова і що кількість результатів для кожного досліджуваного рівня або кількість різних зразків є однаковою, тобто результати є “збалансованими”. Якщо результати “не збалансовані”, то рекомендують результати з відповідних наборів даних відкидати випадковим чином до досягнення “збалансованості”.

При виконанні випробувань на викиди слід розуміти, що викиди не повинні відхилитися лише за статистичними міркуваннями. Для кожної вибірки причину, за якою результат відрізняється від всіх інших, слід досліджувати і ідентифікувати. Випробування на викиди, засноване на використовуваних припущеннях, вказують на наявність достатньої статистичної причини викидів, але вони не вказують, чому це сталося. Лише після проведення повного дослідження для ідентифікації вірогідної причини дані можуть бути оголошені викидами і відкинуті [3].

Якщо конкретний досліджуваний рівень проаналізований з використанням критеріїв Грабса, Кохрена, або інших критеріїв і жодних викидів не виявлено, або вони ідентифіковані, то проводять перевірку інших досліджуваних рівнів. Якщо декілька викидів ідентифіковано в

різних досліджуваних рівнях для конкретного набору даних, отриманих однією лабораторією, необхідно розглянути питання про випробування всіх наборів даних для всіх досліджуваних рівнів.

Крім того, слід вирішувати: необхідно виключати лише викиди, ідентифіковані для цієї лабораторії чи необхідно відхиляти весь набір даних цієї лабораторії. Необхідні дії визначають індивідуально на основі досвіду і знання досліджень, що виконуються для ідентифікації можливих причин.

Для кожної лабораторії чи досліджуваного рівня, або конкретної вибірки при виконанні більшості випробувань на викиди порівнюють деяку міру відносного відхилення підозрілого результату від середнього всіх результатів і оцінюють, чи можливо вважати походження цього результату випадковим.

Критерій Кохрена використовують для перевірки великих стандартних відхилень базових елементів, які збільшують оцінку стандартного відхилення повторюваності, якщо їх враховувати. Статистика, що використовується в критерії Кохрена, тісно пов'язана з h -статистикою Мандела [3].

Критерій Кохрена призначений для обробки внутрішньолабораторних розходжень результатів вимірювань і повинен застосовуватися в першу чергу, після чого повинні бути вжиті корегувальні заходи, у випадку необхідності, з повторенням вимірювань (випробувань).

Для сукупності з p стандартних відхилень s_i , розрахованих виходячи з однієї і тієї ж кількості (n) результатів випробувань в базових елементах, тестова статистика Кохрена має вигляд [8]

$$C = s_{\max}^2 / \sum_{i=1}^p s_i^2, \quad (1)$$

де s_{\max} – найвище значення стандартного відхилення у сукупності.

У випадку, якщо значення тестової статистики менше (чи дорівнює) 5 % критичному значенню, позицію, що випробують, визнають коректною; у випадку, якщо значення тестової статистики більше 5 % критичного значення і менше (чи дорівнює) 1 % значення, позицію, що випробують, називають квазівикидом і помічають однією зірочкою; у випадку, якщо значення тестової статистики більше 1 % критичного значення, позицію, що випробують, називають статистичним викидом і помічають двома зірочками.

Критичні значення для критерію Кохрена представляються у табличній формі і наведені, зокрема у [8].

Однак, критерій Кохрена застосовують у випадках, коли всі стандартні відхилення виходять з однієї і тієї ж кількості (n) результатів вимірювань, отриманих в умовах повторюваності. В фактичних випадках ця кількість може змінюватися за рахунок недостатніх чи виключених даних.

У стандарті ISO 5725-2 тим не менше передбачається, що такі зміни у кількості результатів вимірювань з розрахунку на базовий елемент будуть обмежені і ними можна знехтувати, тобто критерій Кохрена можна використовувати до кількості результатів вимірювань n , що має місце у більшості базових елементів [4].

За допомогою критерію Кохрена перевіряють тільки найвище значення у сукупності стандартних відхилень, і тому така перевірка є односторонньою. Розкид у дисперсіях може також проявлятися у найнижчих значеннях стандартних відхилень. Однак, на малі значення стандартного відхилення може здійснювати дуже сильний вплив ступінь округлення вихідних даних, і тому вони не дуже надійні. Крім того, недоцільно відкидати дані лабораторії через те, що нею досягнута більш висока прецизійність в результатах вимірювань у порівнянні з іншими лабораторіями. Тому критерій Кохрена вважають адекватним.

Якщо найбільше значення стандартного відхилення класифіковане як викид, то воно повинно бути виключене, а перевірка з використанням критерію Кохрена може бути повторена на значеннях, що залишилися. Слід зазначити, що процедура повторення може привести до занадто великих виключень даних у випадку, коли нормальний розподіл, прийнятий за основу, не є достатньо доброю апроксимацією.

Повторне застосування критерію Кохрена пропонується тут лише як корисний засіб, зважаючи на відсутність статистичного критерію, розробленого для перевірки декількох викидів разом. Цей критерій не розроблявся для цієї мети, і висновки при його повторному застосуванні необхідно робити з великою обережністю.

Також обережно треба використовувати критерій Кохрена у випадках, коли результати, що характеризуються великими значеннями стандартних відхилень (особливо якщо вони мають місце у межах лише одного з рівнів), представлені двома чи трьома лабораторіями. З

іншого боку, якщо на різних рівнях у межах однієї лабораторії виявляється декілька квазівикидів і/чи статистичних викидів, то це може бути вагомою вказівкою на те, що внутрішньолабораторна дисперсія занадто велика, і дані цієї лабораторії повинні бути повністю виключені.

Для перевірки, чи не є викидом найбільша величина з x розташованих у порядку зростання сукупності даних x_i ($i = 1, 2, \dots, p$) обчислюють статистику Грабса G_p за виразом [8]

$$G_p = (x_p - \bar{x})/s. \quad (2)$$

де

$$\bar{x} = \frac{1}{p} \sum_{i=1}^p x_i, \quad (3)$$

$$s = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (x_i - \bar{x})^2}. \quad (4)$$

Для перевірки значимості найменшого результату спостережень обчислюють тестову статистику

$$G_1 = (\bar{x} - x_1)/s. \quad (5)$$

У випадку, якщо значення тестової статистики менше (чи дорівнює) 5 % критичного значення, позицію, що випробують, визнають коректною; у випадку, якщо значення тестової статистики більше 5 % критичного значення і менше (чи дорівнює) 1 % значення, позицію, що випробують, називають квазівикидом і помічають однією зірочкою; у випадку, якщо значення тестової статистики більше 1 % критичного значення, позицію, що випробують, називають статистичним викидом і помічають двома зірочками.

Для того, щоб перевірити, чи можуть два найбільших результати спостережень бути викидами, також обчислюють статистику Грабса

$$G = s_{p-1,p}^2 / s_0^2, \quad (6)$$

де

$$s_0^2 = \sum_{i=1}^p (x_i - \bar{x})^2, \quad (7)$$

$$s_{p-1,p}^2 = \sum_{i=1}^{p-2} (x_i - \bar{x}_{p-1,p})^2. \quad (8)$$

а також

$$\bar{x}_{p-1,p} = \frac{1}{p-2} \sum_{i=1}^{p-2} x_i. \quad (9)$$

Відповідно, щоб перевірити два найменших результати спостережень, обчислюють статистику Грабса

$$G = s_{1,2}^2 / s_0^2, \quad (10)$$

де

$$s_{1,2}^2 = \sum_{i=3}^p (x_i - \bar{x}_{1,2})^2, \quad (11)$$

$$\bar{x}_{1,2} = \frac{1}{p-2} \sum_{i=3}^p x_i. \quad (12)$$

Критичні значення для критерію Грабса представляються у табличній формі і наведені, зокрема у [8].

При аналізі експерименту з оцінки прецизійності критерій Грабса може бути застосований до наступних випадків. Аналіз середніх значень базових елементів для заданого.

Спочатку до середніх значень базових елементів рівня j застосовують критерій Грабса для одного викиду. Якщо виявляється, що середнє значення базового елемента є викидом, необхідно виключити його і повторити перевірку для іншого екстремального середнього значення базового елемента (якщо найбільше значення є викидом, то тоді слід перевірити найменше значення, а найбільше значення при цьому виключити), однак при цьому не слід застосовувати критерій Грабса для двох викидів. Цей критерій потрібно застосовувати у випадку, якщо при перевірці з використанням критерію Грабса для одного викиду виявиться, що середні значення базових елементів не мають викидів.

Аналіз вихідних даних в межах базового елемента, для якого в результаті перевірки з використанням критерію Кохрена виявиться сумнівність значення стандартного відхилення.

Критерій Кохрена допускають використовувати для визначення сумісності невеликої дисперсії для конкретного набору даних. Цей критерій ідентифікує дисперсії, які більше середніх дисперсій для цього рівня. В цьому відношенні критерій є одностороннім, оскільки лабораторія з найменшою дисперсією (відносно інших лабораторій) не піддається контролю за цим критерієм.

Рішення про повторення випробувань за критерієм Кохрена залежить від того, чи ідентифіковані викиди, і від кількості наборів даних, що підлягають випробуванню, на конкретному досліджуваному рівні. Якщо викиди не виявлені, то критерій не повторюють, а якщо виявлені, то для перевірки повторюють випробування за критерієм Кохрена на наборах даних, що залишилися, рівня, що розглядається. Ця обережність особливо доречна для невеликої кількості даних, особливо якщо приблизно 20 % даних відхилені як викиди.

Альтернативою випробувань за критерієм Кохрена є випробування за критеріями Бартлета, Левіна і Хартлея. Проте можуть бути випадки, коли викиди ідентифікуються з використанням одного критерію, але не ідентифікуються при використанні іншого критерію.

Критерій Грабса використовують для перевірки наявності базових елементів, які є дуже великими або дуже маленькими і можуть дати найбільшу оцінку стандартного відхилення відтворюваності. Цей критерій головним чином призначений для обробки міжлабораторних розходжень, а також може використовуватися (якщо $n > 2$) у випадках, коли перевірка із застосуванням критерію Кохрена викликала підозру в тому, що висока внутрішньолабораторна варіація зумовлена тільки одним з результатів вимірювань у базовому елементі.

Після застосування критерію Кохрена середні для кожного конкретного досліджуваного рівня, розташовують в порядку зростання. Потім виконують декілька випробувань за критерієм Грабса: спочатку виконують випробування для визначення, чи є найбільше або найменше середнє єдиним викидом, а якщо викид ідентифікований, то його відкидають і повторюють випробування для наступного підозрілого значення.

Для конкретного досліджуваного рівня критерій Грабса для одного викиду дозволяє обчислювати відношення різниці підозрюваного значення і середнього всіх значень рівня до стандартного відхилення всіх значень. Це відношення потім порівнюють з обчисленим або табличним критичним значенням відношення з 95 % і 99 % довірчою ймовірністю.

Якщо одиничний викид не виявлений, виконують випробування за критерієм Грабса для визначення присутності (або відсутності) двох викидів. Випробують два найменших середніх і, якщо викиди не виявлені, то випробують два найбільших середніх. У цьому критерії, якщо обчислене відношення більше розрахункового (або табличного) для заданої довірчої ймовірності, середнє розцінюють як задовільне.

Альтернативою критерію Грабса є критерій Діксона. Проте можливі ситуації, коли викид ідентифікується за одним критерієм, не ідентифікується за іншим.

У алгоритмі, приведеному на рис. 5, вказані основні процедури, які повинні бути виконані при оцінці викидів у отриманих результатах міжлабораторних випробувань. Позначення s_r –

це стандартне відхилення збіжності, а s_R – стандартне відхилення відтворюваності.

Міжнародні стандарти ISO/TR 22971, ISO 5725-2, ISO 13528 і ISO 5479, які застосовуються при статистичному обробленні викидів у отриманих результатах міжлабораторних випробувань розроблені міжнародним технічним комітетом (ТК) Міжнародної організації стандартизації (ISO) ТК 69 “Застосування статистичних методів”. Вони використовуються ISO/ТК, які розробляють проекти міжнародних стандартів у сфері метрології.

Висновки

При обробленні результатів міжлабораторних випробувань, результатів вимірювань, що проводяться лабораторіями, використовуються різноманітні стандартизовані статистичні методи для виявлення викидів у отриманих результатах. При цьому актуальним є розроблення спеціальних рекомендацій щодо особливостей застосування цих методів.

Порівняльний аналіз застосовуваних діаграм і критеріїв для ідентифікації потенційних викидів у отриманих результатах дозволяє обрати саме такі діаграми і критерії, які дозволяють більш ефективно виявляти реальні викиди і проводити більш точне оброблення отриманих результатів.

Список використаних джерел

1. Velychko O., Gordiyenko T. The implementation of general guides and standards on regional level in the field of metrology // Journal of Physics: Conf. Series. – Vol. 238. – 2010. – Numb. 1. – 012044. – 6 p.
2. Величко О. Н., Гордиенко Т. Б. Внедрение международных и региональных нормативных документов в области метрологии на национальном уровне // Системи обробки інформації. – Харків. – 2011. – Вип. 6 (96). – С. 2–8.
3. ISO 5725-2:1994. Accuracy (trueness and precision) of measurement methods and results. – Basic methods for the of repeatability and reproducibility of a standard measurement method.
4. ISO 13528:2005. Statistical methods for use in proficiency testing by interlaboratory comparisons.
5. ISO/IEC 17043:2010. Conformity assessment. – General requirements for proficiency testing.

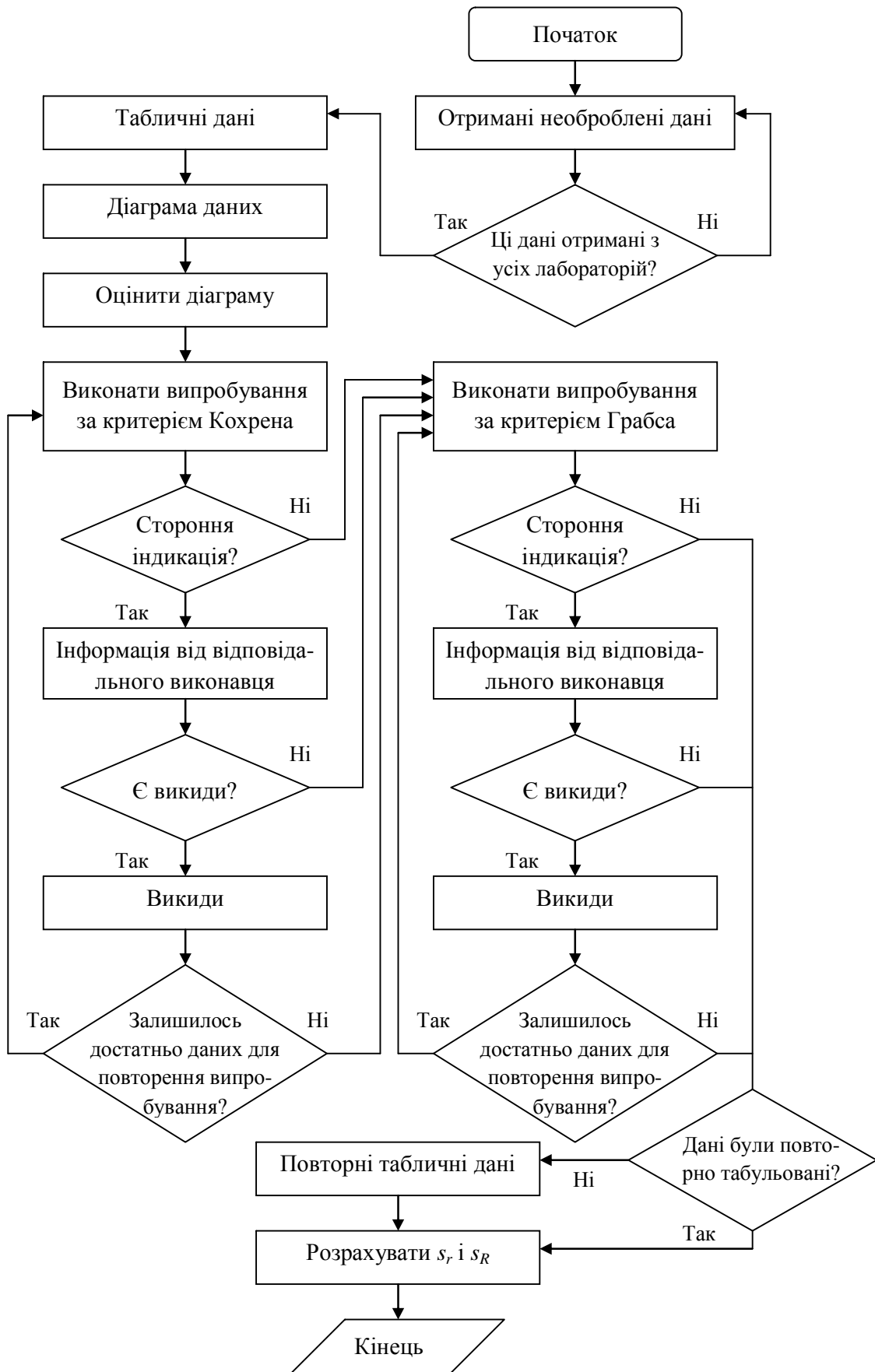


Рисунок 5 – Алгоритм статистичного оброблення викидів у отриманих результатах міжлабораторних випробувань

6. ISO 5479:1997. Statistical interpretation of data. – Tests for departure from the normal distribution.

7. Величко О. М., Коломієць Л. В., Гордієнко Т. Б. Системи управління підприємств і організацій: основні елементи та практична реалізація. Підручник. – Одеса: ВМВ, 2011. – 342 с.

Надійшла до редакції 20.05.2013

Рецензент: д.т.н., доцент Боряк К.Ф.,
Одеська державна академія технічного
регулювання та якості, м. Одеса.

О. Н. Величко., д. т. н., Л. В. Коломиец, д. т. н., Т. Б. Гордиенко, к. т. н.

ПРОВЕРКА КРИТИЧНОСТИ ДАННЫХ МЕЖЛАБОРАТОРНЫХ ИСПЫТАНИЙ С ПРИМЕНЕНИЕМ СТАТИСТИЧЕСКИХ МЕТОДОВ

В международных стандартах стандартизованы разнообразные статистические методы, в частности графические, которые применяются для проведения проверки критичности данных межлабораторных испытаний. Проведен сравнительный анализ стандартизованных графических методов и критериев для выявления выбросов в полученных результатах испытаний. Рассмотрены особенности применения статистических методов и испытаний на выбросы в результатах испытаний.

Ключевые слова: статистический метод, проверка критичности данных, межлабораторные испытания, стандарт.

О. М. Velychko, DSc, L. V. Kolomiets, DSc, T. B. Gordiyenko, PhD

CRITICISM DATA VERIFICATION FOR THE INTERLABORATORY TESTS WITH USING OF STATISTICAL METHODS

Various statistical methods, in particular graphic that apply for criticism data verifications of interlaboratory tests, are standardized in international standards. The comparative analysis of the standardized graphic methods and criteria is realized for the exposure of outlier in the got test results. The features of application of statistical methods and tests on outlier in the test results are considered.

Keywords: statistical method, criticism data verification, interlaboratory test, standard.