**УДК 519. 24 : 543**
**UDC 519. 24 : 543**
**DOI: 10.33744/0365-8171-2023-113.2-117-132**

# НОВА КОНЦЕПЦІЯ ХЕМОМЕТРИКИ

# NEW CONCEPT OF CHEMOMETRICS

***Артеменко Владислав Андрійович, UT5UDJ,*** *магістр екології, Український гідрометеорологічний інститут Державної служби України з надзвичайних ситуацій та Національної академії наук України, науковий співробітник відділу гідрохімії, e-mail: artemenko@uhmi.org.ua, тел. 380936011250, Україна, 03028, м. Київ, просп. Науки 37, к.34.*

*https://orcid.org/0000-0003-0536-5415*

***Петрович Володимир Васильович,*** *кандидат технічних наук, професор, старший науковий співробітник, професор кафедри транспортного будівництва та управління майном Національного транспортного університету. e-mail: petrovichvv60@ukr.net, тел. +380442807338, Україна, 01010, м. Київ, вул. М. Омеляновича-Павленка, 1, к. 138.,*

*https://orcid.org/0000-0003-0422-2535*

**Анотація.** В хемометриці для обробки даних найбільш часто використовуються параметричні статистичні методи досліджень. Але при цьому ігнорується адекватна обробка даних, які нормальному закону розподілення не підлягають. На сучасному етапі широко використовують і класичні непараметричні методи, які не оперують із ймовірнісним розподіленням вихідних даних.

Однак при проведенні безпосередніх практичних розрахунків певні типи розподілень все рівно використовуються.

При реалізації класичних непараметричних методів звичайно необхідне також звернення до відповідних статистичних таблиць.

Для кожного такого методу використовується своя таблиця. Таким чином, хемометрика зараз представлена багатьма не пов'язаними між собою статистичними методами досліджень.

Одним із головних недоліків класичних методів хемометрики (параметричних та непараметричних) є відсутність єдиного підходу до обробки вихідних даних.

У роботі запропонований принципово новий підхід до проведення статистичних досліджень даних.

Використовуючи непараметричний бутстреп, багатий арсенал не пов’язаних між собою методів класичної хемометрики замінюється одним-двома методами без будь-яких принципових обмежень.

З точки зору програмної реалізації це означає наявність універсальних процедур, за допомогою яких можливо вирішувати практично всі задачі хемометрики.

Розглянута одна із важливих задач непараметричного бутстрепу – “розмноження” вибірки.

При цьому імітується, що була проведена не одна серія експериментів, а значно більше (наприклад, 1000 … 10000 серій).

Із використанням метода непараметричного бутстрепа показані особливості визначення довірчого інтервалу для середнього та медіани часового ряду кисню, щобув розчинений у річковій воді.

Приведена програмна реалізація бутстреп-процедур на мові MATLAB, що була розроблена авторами статті. Процедури були реалізовані також на мовах програмування FORTRAN, PASCAL та R.

**Ключові слова:** хемометрика, нова парадигма хемометрики, хемометричні дані, непараметричний бутстреп, універсальні методи та процедури, код MATLAB.

**Мета роботи.** Метою роботи є розробка принципово нової ідейної основи (парадигми) хемометрики, де значна кількість методів класичної хемометрики замінюється лише кількома універсальними методами (універсальними процедурами).

Відповідно, статистична обробка даних повинна бути представлена із єдиних позицій, поза залежністю від конкретних задач та типів розподілень вихідних даних.

*Закон не може бути точним хоча б тому, що поняття, за допомогою яких ми його формулюємо, можуть розвиватись, та в майбутньому виявитись неточними.*

***Альберт Ейнштейн***

*Не шукай помилку. Шукай, як її виправити*

***Генрі Форд***

## Вступ

Як відомо, значна частка досліджень у сучасній екології – це оцінка стану водних об'єктів**.**

Звичайно при цьому вирішуються наступні задачі:

1. Визначення концентрацій забруднюючих речовин у водному середовищі;
2. Обробка даних натурних (польових) досліджень;
3. Отримання узагальнюючих висновків про стан об'єктів, що досліджуються;
4. Прогнозування екологічного стану водних об'єктів на конкретну перспективу.

При виконанні більшості пунктів зазначеного вище переліку задіяні такі дисципліни як аналітична хімія та хемометрика [1-4].

Авторами були розглянуті публікації переважно за період після 2000 року.

Хемометрика – хімічна дисципліна, що застосовує методи математичної статистики для обробки експериментальних даних та видобування при цьому певної корисної інформації.

Базуючись на основоположних методах математичної статистики, хемометрика у значній мірі успадкувала її методологію та термінологію.

Але безпосередньо були також враховані "тонкі" особливості самої хімії, у першу чергу хімії аналітичної. Однак, незважаючи на істотні досягнення у сучасній математичній статистиці, класична хемометрика продовжує базуватись на параметричних методах дослідження, які у більшості своїй вимагають виконання нормального закону розподілення вихідних даних.

Необґрунтоване використання параметричних методів, тобто без проведення статистичних тестів на нормальність та подальшого аналізу результатів цих тестів, стає особливо небезпечним, якщо річ заходить про нормативні документи.

Зважаючи на загальні позиції, річ може іти також про методику адекватної роботи із буть-якими даними, поза залежністю від їх походження.

У цьому зв'язку слід переглянути не тільки методологію, але також практично всю концепцію досліджень у царині хемометрики.

У роботі розглянуті переважно питання хемометрики, безпосередньо пов'язані із обробкою даних та прийняттям відповідних рішень.

Авторами запропонована кардинально нова концепція хемометрики, що використовує принципово тільки один універсальний метод обробки даних – метод непараметричного бутстрепу.

Практично всі задачі хемометрики стає можливим вирішувати за допомогою тільки цього метода.

Параметричні та навіть класичні непараметричні методи слід у кращому разі залишити як допоміжні з метою подальшої порівняльної оцінки роботи різних статистичних процедур.

Хемометрика у цьому разі одержує нову концепцію (нову парадигму).

## Непараметричний бутстреп – максимально можливе втілення непараметричної статистики

На сучасному етапі проведення досліджень у достатній мірі розвинуті та широко використовуються непараметричні методи статистики.

Позитивна якість непараметричних методів полягає в тому, що за своєю суттю вони розраховані на роботу із різноманітними неідеальними даними.

При реалізації цих методів не потрібні спеціальні дослідження, аби встановлювати закони розподілення даних.

У більшості випадків знання конкретного закону взагалі не вимагається!

Однак класичні непараметричні методи мають свій недолік – для кожної окремої статистики потрібно мати свою методику розрахунку.

При цьому такі методи зазвичай не пов'язані між собою.

У цьому зв'язку з позицій еколога – практика бажано мати мінімум методів, які єдиним способом (за єдиною схемою) дозволять обробляти такі дані.

Крім того, враховуючи складність виконання хімічних аналізів, ці методи повинні також адекватно працювати із досить малими об'ємами даних.

З точки зору сучасного рівня знань, ці побажання у повній мірі реалізуються у такому методі математичної статистики, як бутстреп [5-6].

Зважаючи на багатогранність застосування та достатню простоту бутстрепу, кількість публікацій по цьому питанню постійно зростає.

Як відомо, розрізняють бутстреп параметричний та бутстреп непараметричний.

Всі методи бутстрепа, які на тому або іншому етапі проведення досліджень вимагають знання певних законів розподілення, відносять до параметричного бутстрепа.

Методи бутстрепа, які не вимагають знання законів розподілення – до бутстрепу непараметричного.

Далі будемо аналізувати результати застосування тільки бутстрепу непараметричного.

Метод непараметричного бутстрепу дозволяє проводити різноманітні дослідження, одержуючи при цьому числові значення певної статистики.

За допомогою цього метода визначають довірчі інтервали статистик (у тому числі довірчі інтервали коефіцієнтів регресій, кореляцій, тощо), вирішують складні задачі порівняння декількох груп даних, та інше.

В Україні перша публікація по застосуванню непараметричного бутстрепу в хемометриці (гідрохімії) датована 2019 роком [7].

У попередній публікації авторів зазначалось, що вихідні дані слід розглядати із позицій часових рядів, для яких вкрай важливим є порядок окремих складових (точок) цього ряду. Але якщо для часового ряду визначають середнє, медіану, квантиль та інші подібні статистики, то абсолютно неважливим є порядок слідкування складових цього ряду. Тобто такі статистики ніяк не змінять своїх числових значень, якщо навіть складові ряду переставляти у довільному порядку.

Для статистик, що вкрай чутливі до порядку слідкування складових часового ряду, існує окремий клас задач, які у даній статті не розглядаються.

**Розмноження вибірки – базовий принцип непараметричного бутстрепу**

Як відомо, проведення хімічних аналізів являє собою трудомістку процедуру, не кажучи навіть про фінансові витрати на їх виконання.

Тому звичайно обмежуються тільки однією серією експериментів (дослідів). При цьому у такій серії міститься не більше 5 визначень конкретної речовини.

В той же час при класичних методах дослідження (як параметричних, так і непараметричних) вимагається провести 30 … 100 вимірювань в серії.

Зважаючи на цей факт, однією із основних задач непараметричного бутстрепу є "розмноження" вибірки. Тобто імітується, що була проведена не одна серія експериментів, а, наприклад, 5000 або навіть більше.

Будемо надалі оперувати обмеженою вибіркою певного часового ряду, а саму вибірку розглядати у якості популяції.

При цьому кожна підвибірка – це випадкова комбінація із набору елементів вибірки.

Підвибірки повинні мати той самий розмір, що і вихідна вибірка. Для кожної підвибірки можливо знайти числове значення даної статистики.

При генерації підвибірок вирішується задача "розмноження" вихідної вибірки : реалізується алгоритм "Випадковий вибір із зворотом" ("Random sampling with replacement"). Тобто видобуваний елемент може бути повернутий у вибірку та мати шанс бути вибраним знову, тоді як інші елементи можуть бути взагалі відсутні.

Таким чином, призначення непараметричного бутстрепу полягає в тому, щоб випадковим чином багатократно "витягувати" підвибірки із емпіричного розподілення (вибірки), та надалі проводити відповідний статистичний аналіз.

Завдяки випадковому процесу формування підвибірок в результаті одержуємо деяку варіацію числових значень статистики.

Тобто маємо бутстреп – розподілення розкиду числових (точкових) значень.

Зрозуміло, що працюючи із великим масивом даних (звичайно це часові ряди довжиною декілька сотень або навіть більше значень), бутстрен – метод буде повністю орієнтований на застосування ЕОМ.

Для кращого розуміння основоположних принципів цього методу розглянемо простий приклад, що включає тільки 10 бутстреп – випробовувань.

Нехай вихідні дані містяться у векторі

$$V=[1.1;\ 2.2;\ 3.3;\ 4.4;\ 5.5].$$

Необхідно знайти медіану даних та, відповідно, 95% - ні довірчі інтервали для медіани.

Як відомо, у більшості сучасних мов програмування вектор можливо індексувати як за допомогою одного числа, так і за допомогою індексного вектора із цілісночисельними елементами.

При індексуванні вектора $V$ за допомогою одного числа одержуємо також одне число, наприклад,

$$V[1]=1.1;\ \ V[3]=3.3 \text{ і т.д.}$$

При індексуванні цілісночисельним вектором індексний вектор *Indx* буде включати цілі числа від 1 і до довжини вектора. У нашому прикладі це числа 1;2;3;4;5.

Далі будемо застосовувати індексацію вектора тільки цим методом.

Формувати індексний вектор будемо за допомогою генератора випадкових рівномірно розподілених цілих чисел.

При цьому такий генератор повинен також генерувати числа в зазначеному вище діапазоні.

Результати індексування вектора $V$ на кожному кроці бутстреп-іспиту будемо зберігати у векторі *BS_Data.* Зрозуміло, що розмір вектора *BS_Data* повинен бути як і у вектора вихідних даних $V$. Створимо також вектор *BS_Vector.*

У векторі *BS_Vector* будемо зберігати значення статистики (у нашому випадку це медіана), що були одержані при кожному бутстреп-іспиті.

У даному прикладі маємо 10 іспитів.

Внаслідок роботи генератора випадкових чисел при кожному бутстреп-іспиті будемо отримувати відповідно різні значення.

Тому перед проведенням бутстреп-іспитів слід встановити генератор у початковий стан (звичайно це буде "нуль").

Далі наводимо одержані у нашому (*!*) випадку результати бутстреп-іспитів.

Так, для першого бутстреп-іспиту були отримані такі дані:

*Indx=[5; 2; 2; 3; 5];*

*BS_Data= V[Indx] =[5.5; 2.2; 2.2; .3.3; 5.5];*

*Median(BS_Data)=3.3;*

*BS_Vector[1]=3.3;*

для другого бутстреп-іспиту:

*Indx=[2; 5; 5; 4; 4;];*

*BS_Data= V[Indx] =[2.2; 5.5; 5.5; 4.4; 4.4];*
*Median(BS_Data)=4.4;*
*BS_Vector[2]=4.4;*
**……..**
для десятого бутстреп-іспиту
*Indx=[3; 4; 1; 3; 4;];*
*BS_Data= V[Indx] =[3.3; 4.4; 1.1; 3.3; 4.4];*
*Median(BS_Data)=3.3;*
*BS_Vector[10]=3.3.*

В результаті виконання розрахунків отримали значення *BS_Vector* із елементами:

*BS_Vector =[3.3; 4.4; 2.2; 4.4; 4.4; 2.2; 3.3; 4.4; 4.4; 3.3].*

Таким чином, на основі значень бутстреп – розподілення стає можливим визначити довірчі інтервали певної статистики.

**Знаходження границь довірчих інтервалів**

На практиці досить часто зустрічаються задачі знаходження довірчих інтервалів тієї або іншої статистики для конкретного часового ряду.

Однак, зважаючи на те, що числовий ряд може не підкорятись нормальному розподіленню, вирішити такі задачі за допомогою класичних методів практично неможливо. Класичні методи у цьому випадку, наприклад, не дають способів визначення довірчих інтервалів для відносного стандартного відхилення, однак відносне стандартне відхилення широко застосовується на практиці при статистичній обробці експериментальних даних.

При використанні непараметричного бутстрепу знайти довірчі інтервали не представляє складності.

Більш того, для будь-яких статистик та вихідних даних це робиться одностайно.

Розглянемо задачу визначення довірчих інтервалів для ряду розчиненого кисню.

Попередньо необхідно знайти відповідну величину самої статистики.

У нашому прикладі статистика – це медіана, значення *P* довірчої ймовірності дорівнює P=0.95 (або 95%).

Маючи вектор *BS_Vector,* можливо визначити нижню *L* та верхню *H* границі довірчого інтервалу для даної статистики.

Найбільш простий спосіб розрахунку величин *L* та *H* полягає у слідуючому:

1. Задаємося значенням *P;*
2. Знаходимо величину *PL* за формулою *PL=(1-P)/2,*
величину *PH* за формулою *PH=(1+P)/2;*
3. Значення *L* буде дорівнювати

$$L=Quantile(BS_Vector, PL),$$

значення *H* відповідно

$$H=Quantile(BS_Vector, PH).$$

У даному випадку надалі припускається, що *Quantile*( ) – це процедура знаходження квантиля заданого порядку для вектора, який дає той самий результат, що і відповідна функція *MATLAB*.

Скаляри *L* та *H* і будуть рішенням задачі.

Оскільки в прикладі *L*=2.2 (див. вище), а значення *H*=4.4, остаточно одержуємо для медіани 95%-ний довірчий інтервал 2.2 … 4.4.

Зазначимо, що розглянутий простий метод знаходження границь довірчого інтервалу призначений перш за все для використання у разі застосування симетричного розподілення для *BS_Vector* (тобто дає симетричну гістограму).

На практиці цей випадок зустрічається найбільш часто, але не завжди.

Покращений метод розрахунку границь довірчих інтервалів, що адекватно працює у тому числі із несиметричними бутстреп-розподіленнями, визначається згідно слідуючих пунктів:

1. Призначаємо величину *P;*
2. Знаходимо значення *PL* за формулою

$$PL=(1+P)/2,$$

величину *PH* за формулою

$$PH=(1-P)/2;$$

3.Визначаємо проміжну величину *TMP*, яка дорівнює

*TMP*,*=2 · Mean (BS_Vector);*

4.Знаходимо *L* за формулою *L=TMP-Quantile (BS_Vector, PL), величину H* за формулою *H=TMP-Quantile (BS_Vector, PH).*

Визначені скалярні величини і є рішенням задачі.

При цьому сам бутстреп – вектор *(BS_Vector)* формується як у попередньому методі, тобто модифікація не зачіпає саме спосіб його формування.

На практиці краще використовувати другий (покращений) метод, тим більше що він не набагато складніший за метод перший.

Однак за рахунок незначного ускладнення другого метода досягається вагоме покращення якості розрахунків.

Окремо слід зазначити, що оскільки в обох алгоритмах використовується генератор випадкових чисел, у загальному випадку, навіть при однакових вихідних даних, можуть бути одержані різні значення довірчих інтервалів (але достатньо близькі між собою).

З точки зору математичної статистики такі результати будуть цілком коректними.

Аналогічно знаходимо довірчі інтервали при інших значеннях величини *P* та будь-якої статистики, тобто середнього, медіани, квантилів певного порядку, коефіцієнтів варіації, тощо.

Це і є той узагальнюючий принцип бутстреп – методу при визначенні статистик вихідних даних.

Далі коротко зупинимось на окремому питанні, яке необхідне для більш повного висвітлення суті проблеми.

У деяких наукових публікаціях поширюється думка, що довірчі інтервали при використанні непараметричних методів виходять більш широкими, ніж при застосуванні методів параметричних (теж саме при перевірці різних статистичних гіпотез).

Тобто для одержання довірчих інтервалів рівної ширини у разі застосування непараметричних методів нібито необхідний більший об’єм вихідних даних у порівнянні із методами параметричними.

Однак, як показали проведені авторами числові експерименти, такі уявні переваги параметричних методів для бутстрепа на практиці зникають.

**Обговорення кода процедури**

Процедура знаходження точкової та інтервальних оцінок даної статистики носить назву *BSSTAT ( )* . Вона визначає коментарі, достатні для її практичного використання.

У зв’язку з тим, що на практиці користувачі не застраховані від різного роду помилок, пов’язаних перш за все із некоректністю даних, які подають на вхід процедури, приблизно третину коду складають необхідні перевірки.

Процедура *ISRV ( )* перевіряє вхідний аргумент *IN*. Аргумент повинен бути числовим вектором (“горизонтальним” або “вертикальним” за термінологією MATLAB).

Даний аргумент – це матриця розміром *1xSZ або SZx1,* яка не містить комплексних чисел, а також *NAN, -Inf* та *+Inf* значень.

Далі йде процедура *ISFH ( ),* яка перевіряє правомірність вказівника (такого, що вказує тільки на реально існуючу функцію).

Однак така перевірка не зачіпає кількості аргументів процедури, на яку вказує *F*.

У даному випадку це повинна бути процедура, що допускає свій виклик тільки з одним аргументом, та повертає один аргумент у вигляді числа (скаляра). Якщо, наприклад, процедура *quantile ( )* вимагає, як мінімум, двох вхідних аргументів, використання *F=@quantile* у даному контексті неприпустимо.

Далі контролюється, щоб розмір вихідного вектора *IN* складав не менш ніж 3 значення.

Адекватні результати можуть бути одержані вже при довжині вектора *IN* близько 8 значень. В принципі, довжина вектора вихідних даних повинна мати не менше 32 значень.

Під бутстреп – вектор відводиться пам'ять (строка 42 кода) за допомогою *zeros( )*.

Необхідні змінні *INDX* та *V* далі декларуються у вигляді пустих масивів (без такої декларації, в принципі, можливо і обійтись).

Далі в циклі із числом повторень N=10000 (число бутстреп - іспитів) проводяться основні обчислення.

Обчислення включають етапи, що розглянуті вище в тексті статті:

1. Формування індексного вектора *INDX;*
2. Виборка згідно індексів вектора *INDX* із вихідних даних *IN* у вектор *V*. При цьому вектори *INDX* (*IN* та *V*) мають однакову довжину;
3. Знаходження поточної статистики для вектора *V*. Здійснюється за допомогою *F(V)*, тобто результат при *F=@mean* буде таким, як і при безпосередньому виконанні *mean(V)*.

В результаті виконання *F(V)* одержуємо звичайне число (скаляр), тобто числове значення даної статистики.

Зрозуміло, що всі перевірки не має сенсу реалізовувати у повному обсязі, оскільки деякі із помилок зустрічаються вкрай рідко.

Крім того, кожна нова перевірка вхідних аргументів вносить свою частку у сповільнення роботи процедури (особливо це помітно, якщо процедура буде знаходитись всередині циклу).

4. Знайдене числове значення поточної статистики записується у бутстреп-вектор *VBS* за допомогою *VBS(J)=F(V)* згідно строки 47 кода.

Далі знаходимо значення *PL; PH; TMP* та остаточно визначаємо значення *L* та *H*. Результати обчислень записуємо у змінній *OUT*.

*OUT* – це структура із полями *HIGH* (значення верхньої границі 95%-го довірчого інтервалу) та *LOW* (відповідне значення нижньої границі).

Зазначимо, що одне із полів буде ім’ям статистики (ім’ям функції, яка визивається), що знаходиться у верхньому регістрі.

Це поле є точковим значенням цієї статистики.

Поле *CORRECTED* – це скоректоване (за допомогою бутстреп – метода) точкове значення статистики, тобто “середня величина” бутстреп – вектора *VBS*.

У своїй роботі процедура *BSSTAT( )* звертається до інших процедур, коди яких із необхідним мінімумів пояснень наведені на лістингах.

Далі розглянемо приклад, що ілюструє роботу із функціями, які вимагають більше одного вхідного аргумента.

Якщо необхідно знайти, наприклад, довірчі інтервали 75%-го квантиля, слід використати функцію користувача типу

```
function[OUT]= QUANTILEØ75(IN)
 OUT=quantile (IN,0.75);
 end
```

При цьому, очевидно, виклик процедури повинен бути

*F=@QUANTILE Ø75*

Оскільки наведений вище код поданий на стандартній мові *MATLAB*, для роботи у безкоштовних аналогах цієї системи може знадобиться незначна модернізація цього кода.

Відповідні універсальні процедури були реалізовані також на мовах програмування *FORTRAN, PASCAL* та їх діалектах, а також на *R*.

F:\BSSTAT.M
Страница 1 из 2 2023-04-04 23:31:03

```
function [ Out ] = BSSTAT( In , F )

Out = [ ];

%============================================================%
% Finding Value Of The Statistics And Confidential          %
% Interval Of This Statistics At The Method Of              %
% NonParametric BootStrap                                   %
%============================================================%
% In ===> Raw Data ( Vector )                               %
%============================================================%
% F ====> Pointer To Function Which Calculates Statistics   %
% Example : F = @ mean                                      %
%============================================================%

P = 0.95;
N = 10000;

%============================================================%
% P ===> Confidential Probability                           %
%============================================================%
% N ===> Number Of BootStrap Tests                          %
%============================================================%

if not( IsRV( In ) )
 return
end

if not( IsFH( F ) )
 return
end

In = In( : );
SZ = length( In );

if SZ < 3
 return
end

Indx = [ ];
V = [ ];
VBS = zeros( N , 1 );

```

- 1 -

F:\BSSTAT.M
Страница 2 из 2 2023-04-04 23:31:03

```
44  for J = 1 : N
45   Indx = IRand( SZ , 1 , 1 , SZ );
46   V = In( Indx );
47   VBS( J ) = F( V );
48  end
49
50  PL = ( 1 + P ) ./ 2;
51  PH = ( 1 - P ) ./ 2;
52  Tmp = 2 .* mean( VBS );
53  L = Tmp - quantile( VBS , PL );
54  S = F( In );
55  H = Tmp - quantile( VBS , PH );
56  C = mean( VBS );
57  Out.CORRECTED = C;
58  Out.HIGH = H;
59  Out.LOW = L;
60  Out = setfield( Out , upper( func2str( F ) ) , S );
61
62  end
63
```

- 2 -

F:\IRAND.M
Страница 1 из 1 2023-04-04 23:31:28

```
function [ Out ] = IRand( SZ01 , SZ02 , IMin , IMax )

Out = [ ];

%============================================================%
% Generation The Matrix By Size SZ01 X SZ02 Of Integer      %
% Numbers In Range IMin ... IMax With Uniform               %
% Probabilistic Law Of The Distribution                     %
%============================================================%

if not( IsPIN( SZ01 ) )
 return
end

if not( IsPIN( SZ02 ) )
 return
end

if not( IsIN( IMin ) )
 return
end

if not( IsIN( IMax ) )
 return
end

Out = floor .......
( IMin + ( IMax - IMin + 1 ) .* rand( SZ01 , SZ02 ) );

end
```

- 1 -

F:\ISFH.M
Страница 1 из 1 2023-04-04 23:31:44

```
 1  function [ Out ] = IsFH( In )
 2
 3  Out = false;
 4
 5  %=========================================================%
 6  % Is "In" Pointer To Real Existing Function ???           %
 7  %=========================================================%
 8
 9  if not( isa( In , 'function_handle' ) )
10   return
11  end
12
13  if exist( func2str( In ) ) > 0
14   Out = true;
15  end
16
17  end
18
```

F:\ISIN.M
Страница 1 из 1 2023-04-04 23:31:59

```
 1  function [ Out ] = IsIN( In )
 2
 3  Out = false;
 4
 5  %============================================================%
 6  % Is "In" Integer Number ???                                 %
 7  %============================================================%
 8
 9  if isempty( In )
10   return
11  end
12
13  if not( isnumeric( In ) )
14   return
15  end
16
17  if not( isscalar( In ) )
18   return
19  end
20
21  if isnan( In )
22   return
23  end
24
25  if isinf( In )
26   return
27  end
28
29  if not( isreal( In ) )
30   return
31  end
32
33  if In ~= fix( In )
34   return
35  end
36
37  Out = true;
38
39  end
40
```

- 1 -

F:\ISPIN.M
Страница 1 из 1 2023-04-04 23:32:09

```
 1  function [ Out ] = IsPIN( In )
 2
 3  Out = false;
 4
 5  %=========================================================%
 6  % Is "In" Positive Integer Number ???                    %
 7  %=========================================================%
 8
 9  if isempty( In )
10   return
11  end
12
13  if not( isnumeric( In ) )
14   return
15  end
16
17  if not( isscalar( In ) )
18   return
19  end
20
21  if isnan( In )
22   return
23  end
24
25  if isinf( In )
26   return
27  end
28
29  if not( isreal( In ) )
30   return
31  end
32
33  if In ~= fix( In )
34   return
35  end
36
37  if In < 1
38   return
39  end
40
41  Out = true;
42
43  end
```

- 1 -

F:\ISRV.M
Страница 1 из 1 2023-04-04 23:32:32

```
function [ Out ] = IsRV( In )

Out = false;

%=============================================================%
% Is "In" A Real Vector ???                                   %
%=============================================================%

if isempty( In )
 return
end

if not( isnumeric( In ) )
 return
end

if not( isreal( In ) )
 return
end

if not( isvector( In ) )
 return
end

if any( isnan( In ) )
 return
end

if any( isinf( In ) )
 return
end

Out = true;

end
```

- 1 -

**Висновки**

Розглянута неправомочність гіпотези про нормальне розподілення вихідних даних в екології.

На базі сучасних методів статистики запропонована уніфікована методика аналізу таких даних.

Методика максимально використовує метод непараметричного бутстрепа, що дозволяє за єдиною схемою визначати довірчі інтервали будь-яких статистик, виконувати складні екологічні розрахунки поза залежністю від статистичних законів розподілення вихідних даних.

Визначена реальна можливість імітації значного числа серій експериментів, що дозволило одержувати результати, близькі до реальної дійсності.

З метою реалізації відповідної методики розроблені пакети програмного забезпечення (на 4-х мовах програмування).

Нова парадигма, по-суті, не обмежується виключно використанням базових положень хемометрики, а може широко трактуватись як загальний метод обробки даних, що одержані при проведенні натурних експериментів в рамках інших наукових дисциплін.

**Перелік посилань**

1. Meier P.C., Zünd R.E. Statistical Methods in Analytical Chemistry. Second Edition. John Wiley & Sons, Inc. 2000. -448p.
2. Practical Guide to Chemometrics. Second Edition. Edited by P. Gemperline. Published CRC Press, 2006 – 520 p.
3. Тарасова В.В. Екологічна статистика. К.: Центр учбової літератури, 2008. – 392 с.
4. Miller J.N., Miller J.C. Statistics and Chemometrics for Analytical Chemistry. Sixth Edition. Prentice Hall, 2010. – 297 p.
5. Chernick M.R., Labudde R.A. An introduction to bootstrap methods with applications to R. John Wiley & Sons Inc. Hoboken, New Jersey, Canada, 2011. –236p.
6. Zieffler A.S., Harring J.R., Long J.D. Comparing groups : randomization and bootstrap methods using R. John Wiley & Sons Inc. Hoboken, New Jersey, Canada, 2011. – 331 p.
7. Артеменко В.А., Петрович В.В. Сучасні статистичні методи обробки екологічних даних. Автомобільні дороги і дорожнє будівництво, вип. 105. К. : Вид-во НТУ, 2019. –С. 35-43.

## NEW CONCEPT OF CHEMOMETRICS

**Artemenko Vladuslav A.,** ***UT5UDJ,*** Master of Ecology, Ukrainian Hydrometeorological Institute, State Service on Emergencies of Ukraine and National Academy of Science of Ukraine, Hydrochemical Research, Scientific Employee, e-mail: artemenko@uhmi.org.ua, tel. 380936011250, Nauki avenue, 37, Kyiv, Ukraine, 03028, room 34, https://orcid.org/0000-0003-0536-5415
**Petrovych Volodymyr V.,** Candidate of Technical Sciences, Professor, Senior Researcher, Professor of the Transportation Construction and Property Management Department**,** National Transport University. e-mail: petrovichvv60@ukr.net, tel. +380442807338, Ukraine, 01010, Kyiv, street M. Omelyanovicha-Pavlenka, 1, room 138, https://orcid.org/0000-0003-0422-2535

**Abstract.** The parametric statistical methods of the research experimental data most often use presently in chemometrics.

If experimental data do not correspond to the normal probabilistic distribution that in this case it is impossible prodice adequate data processing.

At recently in chemometrics more often started to use the classical nonparametric methods. The classical nonparametric methods do not handle the raw datas probabilistic distribution.

But these methods when undertakind the real calculations don't care use that or other types of the distributions.

The classical nonparametric methods under its realization usually requare referencing to corresponding statistical tables.

Its own table is used for each such method. Chemometrics to presently presents itself many not bound between itself statistical methods of the investigations.

The huge defect of the classical methods of chemometrics (parametric and nonparametric) in absence of the united approach to processing the raw datas.

In this article is offered in fundamentally new approach to statistical investigations of the datas.

If use the method of nonparametric bootstrap that possible replace many unbound between itself methods of classical chemometrics whole one ore two universal methods.

And these universal methods have not what or essential defect.

With standpoint of the program realization this means presence one or two universal procedures for decision nearly all practical problems of chemometrics.

One of the primary tasks of nonparametric bootstrap as follows problem of the duplication of the sample is considered in article.

Herewith simulated that was organized not one series of experiment but well over (for example 1000 … 10000).

On example is shown use the method of nonparametric bootstrap with finding confidential interval for average and median for ecological time series.

On programming language MATLAB is brought code corresponding bootstrap procedure.

**Keywords:** chemometrics, new paradigm of chemometrics, nonparametric bootstrap, real chemometrics data, universal methods and procedures, MATLAB code.

**Aim of the work.** Aim of the work in development in fundamentally new paradigm of chemometrics where large number of methods of classical chemometrics is replaced by several universal methods (universal procedures).

Accordingly now statistical data processing is represented since unit (universal) position outside of dependencies from concrete solved problems and type of the statistical distribution of the raw datas.

## References

1. Meier P.C., Zünd R.E. Statistical Methods in Analytical Chemistry. Second Edition. John Wiley & Sons, Inc. 2000. -448p.
2. Practical Guide to Chemometrics. Second Edition. Edited by P. Gemperline. Published CRC Press, 2006 – 520 p.
3. Tarasova V.V. Ekolohichna statystyka. K.: Tsentr uchbovoyi literatury, 2008. – 392 s.
4. Miller J.N., Miller J.C. Statistics and Chemometrics for Analytical Chemistry. Sixth Edition. Prentice Hall, 2010. – 297 p.
5. Chernick M.R., Labudde R.A. An introduction to bootstrap methods with applications to R. John Wiley & Sons Inc. Hoboken, New Jersey, Canada, 2011. –236p.
6. Zieffler A.S., Harring J.R., Long J.D. Comparing groups : randomization and bootstrap methods using R. John Wiley & Sons Inc. Hoboken, New Jersey, Canada, 2011. – 331 p.
7. Artemenko V.A., Petrovych V.V. Suchasni statystychni metody obrobky ekolohichnykh danykh. Avtomobil′ni dorohy i dorozhnye budivnytstvo, vyp. 105. K. : Vyd-vo NTU, 2019. –S. 35-43.