

УДК 378 (004.42)

DOI 10.5281/zenodo.14567320

В. М. Базурін

ORCID ID 0000-0002-6614-4889

Державний торговельно-економічний університет

С. В. Базуріна

Глухівський міський центр позашкільної освіти

## ПРОГРАМНА РЕАЛІЗАЦІЯ АЛГОРИТМУ РОЗРАХУНКУ КРИТЕРІЮ МАННА-УЇТНІ НА МОВІ C#

*Проведено дослідження методів статистичної обробки результатів експерименту (у тому числі й педагогічного). Показано важливість статистичної обробки результатів експерименту. Особливу увагу приділено критерію Манна-Уїтні. Розкрито алгоритм розрахунку критерію Манна-Уїтні для двох вибірок даних. У статті також наведено блок-схему алгоритму розрахунку критерію Манна-Уїтні для двох вибірок даних. У статті наведено послідовність розробки програми для обчислення критерію Манна-Уїтні для двох вибірок. Використано компонентно-орієнтований підхід і об'єктно-орієнтовану парадигму програмування. Розроблено діаграму класів модуля, який виконує обчислення критерію Манна-Уїтні для двох вибірок з точністю  $\alpha=0,05$ . Створений модуль реалізовано у вигляді DLL-бібліотеки, яка підключається до створюваного додатка, причому додаток можна створювати на різних мовах програмування, які підтримують інтеграцію у проект DLL-бібліотек. У статті також описано поля і методи основних класів створеного модуля, наведено приклади кодів і вхідних даних. Розроблений модуль може бути підключений до програмного засобу, який здійснюватиме статистичну обробку результатів експерименту.*

**Ключові слова:** алгоритм, програмна реалізація, мова програмування, критерій Манна-Уїтні, програмний модуль, DLL-бібліотека.

**Постановка проблеми.** Експеримент є однією з важливих складників наукового дослідження у галузі природничих та соціальних наук. Часто метою експерименту є перевірка обраної методики, визначення відмінностей між двома масивами значень тощо. Тобто, математична обробка результатів експерименту зводиться до порівняння двох або більше наборів даних: контрольної та експериментальної вибірок, експериментальної вибірки до експерименту і тієї ж вибірки після експерименту тощо. У процесі порівняння двох масивів даних використовуються методи математичної статистики. Лише ці методи надають можливість зробити висновок, чи є виявлені відмінності між двома масивами даних значущими.

За минулі 150 років вченими-математиками розроблена значна кількість статистичних методів для оцінювання відмінностей між двома масивами даних. Деякі з них більш точні, деякі – менш точні, деякі придатні для порівняння лише великих за обсягом масивів даних, деякі – для порівняння малих вибірок.

Існує низка програмних засобів, призначених для статистичної обробки результатів експерименту. Проте частина їх – платні, для користування іншими засобами (MatLAB, MathCAD та ін.) потрібно вивчити не лише методи математичної статистики, а й відповідну мову, яка використовується у програмному засобі.

Застосування такого програмного засобу, як MS Excel, також має свої недоліки: потрібно скласти таблицю розрахунків і ввести в документ MS Excel таблицю істинності. Все це значно збільшує час статистичної обробки результатів експерименту і збільшує ймовірність помилки.

Одним з статистичних критеріїв, призначених для порівняння двох невеликих вибірок, є критерій Манна-Уїтні. Проте через деякі особливості цього критерію існують труднощі в його програмній реалізації.

**Аналіз актуальних досліджень.** Проблема, яку ми розв'язуємо у цій статті, стосується двох різних галузей науки – основ наукових досліджень (або математичної статистики) і розробки програмного забезпечення (програмної інженерії). Проаналізуємо наукові праці з цих галузей.

Проблеми застосування статистичних критеріїв в обробці результатів педагогічного експерименту досліджувалися такими вченими, як М.Рубцов, Є.Муртазієв, Н.Рубцова [9], В.Петрук, О.Семеніхіна, Ю.Сабадош [6] та інші.

Зокрема, у дослідженні М.Рубцова, Е.Муртазієва, Н.Рубцової висвітлена методика вибору статистичних критеріїв у процесі психолого-педагогічних досліджень, розкрито основні проблеми, пов'язані з оцінюванням ефективності процесу навчання, наведено приклад обчислення критеріїв у програмі «Педагогічна статистика» [9].

У статті В.Петрук, О.Семеніхіної, Ю.Сабадош [6] наведено приклади застосування статистичних критеріїв для оцінювання результатів педагогічного експерименту і їх реалізацію у середовищі електронних таблиць MS Excel.

Г.Кирилецька пропонує застосовувати для статистичної обробки результатів педагогічного експерименту відповідні програмні засоби: MatLAB, Statistica, SPSS та ін. [3].

Інші дослідження приділяють значну увагу компонентно-орієнтованому підходу до розробки програмного забезпечення.

Компонентно-орієнтований підхід вважається один із сучасних підходів до розроблення програмних продуктів. Проблемам застосування компонентно-орієнтованого підходу у розробці програмних засобів присвячені дослідження О.В.Співаковського, В.С.Круглика, В.М.Грищенко, Т.І.Коробейнікової, С.С.Омельченка, С.Я.Ваврука, В.Б.Копей, В.І.Рибнікова та ін.

Зокрема, у дослідженнях О.В.Співаковського і В.С.Круглика здійснено побудову системи дистанційного навчання з використанням компонентно-орієнтованого підходу [11]. Розроблена науковцями система складається з 3 основних компонентів: компонент для реалізації входу в систему, компонент безпеки і компонент адміністрування. Ці компоненти деталізовано у вигляді менших компонентів з більш вузькими функціями: «Навчальний посібник», «Зошит», «Середовище для розв'язування задач», компонент для проведення моніторингу, компонент для групового обговорення проблем, генератор завдань [11].

У дослідженнях Т.І.Коробейнікової і С.С.Омельченка описано основні компоненти ігрового додатка [5].

С.Я.Ваврук визначає такі модулі мережевого Java-дodatка: стандартний прикладний інтерфейс (ODBC API), спеціальний менеджер (ODBC Driver Manager), драйвер ODBC, провайдер даних та ін. [0].

У статті В.Б.Копей наведено приклад застосування компонентно-орієнтованого підходу до моделювання ланок плоского механізму. Автором розроблено модулі додатка і наведено код цих модулів [4]. Результати дослідження можуть бути використані під час розробки програмних засобів, які моделюють плоскі шарнірні механізми, що вивчаються у курсі технічної механіки вищих закладів освіти.

В.І.Рибніков розглядає особливості застосування компонентно-орієнтованого підходу у проектуванні систем штучного інтелекту [8].

**Мета статті** – розкрити особливості програмної реалізації алгоритму розрахунку критерію Манна-Уїтні для двох вибірок.

**Виклад основного матеріалу.** Розглянемо спочатку особливості розрахунку статистичного критерію Манна-Уїтні.

Критерій Манна-Уїтні – критерій, який призначений для оцінювання різниці між двома вибірками або двома популяціями [13]. Алгоритм обчислення критерію Манна-Уїтні має такий вигляд:

- 1) дані обох вибірок об'єднати в один масив;
- 2) виконати ранжування оцінок і виписати окремо ранги для першої та другої вибірок;
- 3) обчислити суму рангів окремо для кожної вибірки;
- 4) обчислити критерій  $U$  для кожної вибірки:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1 \quad (1)$$

Де  $n_1$  і  $n_2$  – обсяги кожної з вибірок;

$\sum R_1$  – сума рангів першої вибірки.

Для другої вибірки обчислення виконуються за аналогічною формулою:

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2 \quad (2)$$

5) взяти менший з двох критеріїв  $U$  – це і буде критерій Манна-Уїтні.

Далі отримане значення критерію Манна-Уїтні  $U$  порівнюється з критичним при  $\alpha=0,05$  [12; 13]. Якщо  $U < U_{\text{табл}}$ , відмінності значущі. Якщо  $U > U_{\text{табл}}$ , відмінності незначущі.

Під час програмної реалізації алгоритму розрахунку критерію Манна-Уїтні виникають деякі складнощі, а саме:

1) у результаті ранжування обидві вибірки змішуються. Далі складно визначити, якій саме групі належить конкретне значення. Тобто, потрібно писати додаткову функцію для визначення вибірки, якій належить це значення;

2) подання критичних значень у вигляді таблиці, де кількість стовпчиків і рядків дорівнює 20.

Для розв'язання першої проблеми програма повинна оперувати не з числами вибірок, а з екземплярами класу, кожен з яких характеризується значенням, рангом і назвою вибірки. Після цього досить просто відсортувати всі об'єкти за спаданням і обчислити суму рангів для кожної вибірки окремо.

Блок-схему розрахунку критерію Манна-Уїтні наведено на рис. 1.

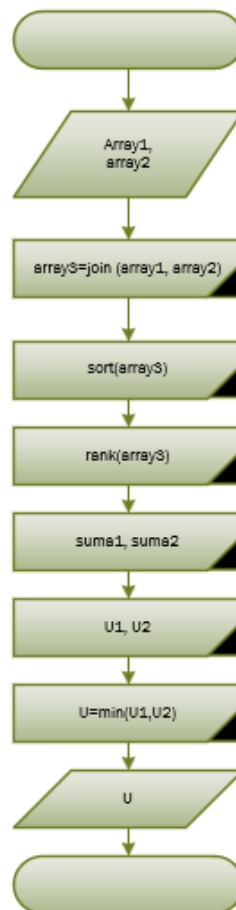


Рис. 1. Блок-схема алгоритму розрахунку критерію Манна-Уїтні.

Під час програмної реалізації алгоритму розрахунку критерію Манна-Уїтні будемо застосовувати компонентно-орієнтований підхід і об'єктно-орієнтовану парадигму програмування.

Компонентно-орієнтований підхід передбачає модульну структуру додатка, створення його з окремих компонентів. Кожен компонент обробляється і тестується

окремо. У даному випадку ми розробляємо модуль, який буде обчислювати критерій Манна-Уїтні і робити висновок про значущість виявлених відмінностей між вибірками. Кожен такий модуль компонується у вигляді бібліотеки класів і функціонує незалежно від інших. Даний модуль компілюється в DLL-бібліотеку і може бути підключений до іншого додатка як незалежний компонент.

Під час розробки модуля було застосовано об'єктно-орієнтовану парадигму програмування. Відповідно до неї під час розрахунків ми будемо оперувати з об'єктами певного класу.

У результаті розробки було створено класи NumberValue і MannWhitney (рис. 2).

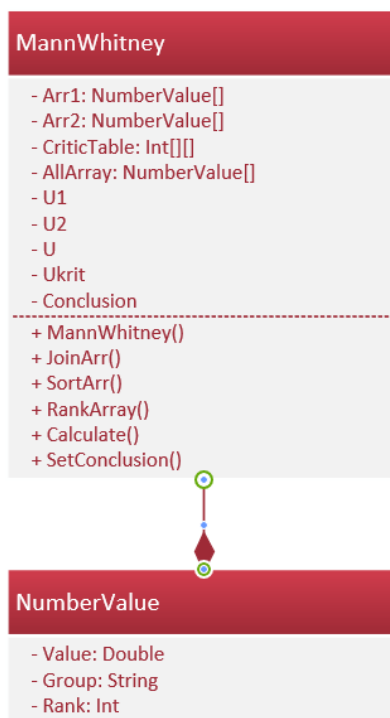


Рис. 2. Діаграма класів.

Клас NumberValue описує окреме значення  
Опишемо основні поля класу NumberValue:

- Value – числове значення;
- Group – група (назва або номер групи);
- Rank – ранг у загальному масиві.

Клас MannWhitney виконує основні обчислення. Цей клас містить такі поля:

- Arr1 – перша вибірка;
- Arr2 – друга вибірка;
- CriticTable – таблиця критичних значень критерію Манна-Уїтні;
- AllArray – об'єднаний масив значень;
- U1 – критерій Манна-Уїтні для першої вибірки;
- U2 – критерій Манна-Уїтні для другої вибірки;
- U – обчислене значення критерію Манна-Уїтні для обох вибірок;
- Ukrit – критичне значення критерію Манна-Уїтні;
- Conclusion – висновок про подібність / відмінність вибірок.

Клас MannWhitney має такі методи:

- MannWhitney() – конструктор класу за замовчуванням;
- JoinArr() – об'єднання двох масивів в один;
- SortArr() – сортування об'єданого масиву за спаданням;
- RankArray() – ранжування елементів об'єданого масиву;
- Calculate() – обчислення критерію Манна-Уїтні;
- SetConclusion () – формулювання висновку.

## Код класу NumberValue

```

public class NumberValue
{
    double value;
    int rank;
    string group;
    public xData(){
        val=0;
        rank=0;
        group="";
    }
    public NumberValue (double x, string g)
    {
        val=x;
        group=g;
        rank=0;
    }
    public double Value{
        get{
            return(val);
        }
    }
    public int Rank{
        get{
            return rank;
        }
        set{
            rank=value;
        }
    }
    public string Group{
        get{
            return group;
        }
    }
}

```

## Код метода для обчислення критерію Манна-Уїтні

```

public double calculate(double[] eGroup, double[] cGroup){
    int N1=eGroup.Length;
    int N2=cGroup.Length;
    int i,j;
    if(N1<21&&N2<21){
        pMin=p1[N1,N2];
        pMax=p5[N1,N2];
        MessageBox.Show(Convert.ToString(pLeft));
    }
    else{
        MessageBox.Show("the groups are too large");
    }
    double x;

```

```

xData obj;
xData[] arr1=new xData[N1];
xData[] arr2=new xData[N2];
int N=N1+N2;
for(i=0;i<N1;i++){
    x=eGroup[i];
    arr1[i]=new xData(x,"e");
}
for(i=0;i<N2;i++){
    x=cGroup[i];
    arr2[i]=new xData(x,"k");
}
xData []rankArr=new xData[N1+N2];
for(i=0;i<N1;i++){
    obj=arr1[i];
    rankArr[i]=obj;
}
for(i=0;i<N2;i++){
    obj=arr2[i];
    rankArr[i+N1]=obj;
}

double x1,x2;

for(i=0;i<N;i++){
    for(j=i+1;j<N;j++){
        x1=rankArr[i].Value;
        x2=rankArr[j].Value;
        if(x1>x2){
            obj=rankArr[i];
            rankArr[i]=rankArr[j];
            rankArr[j]=obj;
        }
    }
}
StreamWriter sw=new StreamWriter("masiv.txt");
for(i=0;i<N;i++){
    rankArr[i].Rank=i+1;
    string ss=Convert.ToString(rankArr[i].Value)+"
"+Convert.ToString(rankArr[i].Rank)+"==" +rankArr[i].Group;
    sw.WriteLine(ss);
}
sw.Close();
int nx;
int Tk=0, Te=0, Tx=0, cur;
for(i=0;i<N;i++) {
    cur=rankArr[i].Rank;
    if(rankArr[i].Group=="e"){
        Te+=cur;
    }
    if(rankArr[i].Group=="k"){
        Tk+=cur;
    }
}

```

```

    }
}
if(Tk>Te) {
    Tx=Tk;
    nx=N1;
}
else{
    Tx=Te;
    nx=N2;
}
string s=Convert.ToString(Tk);
s="Tk="+s;
MessageBox.Show(s);
s=Convert.ToString(Te);
s="Te="+s;
MessageBox.Show(s);
s=Convert.ToString(Tx);
s="Tx="+s;
MessageBox.Show(s);

U=N1*N2+(nx*(nx+1))/2-Tx;
s=Convert.ToString(U);
s="U="+s;
MessageBox.Show(s);

return(U);
}

```

Лістинг 3

Код метода для валідації результатів експерименту

```

public string validate(){
    string res="";
    if(U>pMax){
        res="The groups are similar";
    }
    else{
        if(U>pMin){
            res="The result influently";
        }
        else{
            res="Groups differs";
        }
    }
    return(res);
}

```

Формат вхідних даних: екземпляр класу MannWhitney отримує 2 масиви дійсного типу. Ці дані зберігаються у полях класу і всі подальші операції виконуються саме з ними.

Після написання коду дану програму було скомпільовано у файл DLL. Цей файл доцільно використовувати у вигляді окремого компонента. Для цього його потрібно підключити до проекту і вказати простір імен у заголовку програми.

**Висновки та перспективи подальших наукових розвідок.** Під час розробки програми було застосовано компонентно-орієнтований підхід. Це дозволяє підключати

створений модуль у вигляді DLL-бібліотеки і під час розробки проекту зосередити всю увагу над створенням інтерфейсу додатка і налагодження введення і виведення даних.

Розроблений нами компонент було скопільовано у вигляді DLL-бібліотеки. У подальшому його було інтегровано в додаток, який здійснює статистичну обробку результатів експерименту за кількома критеріями (Пірсона, Вілкоксона, Фішера, Ст'юдента).

#### **СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ / REFERENCES**

1. Ваврук, Є. Я. (2017). Компонентно-орієнтований підхід в програмуванні java-додатків. Нафтогазова енергетика–2017. Матеріали VI Міжнародної науково-технічної конференції (16 травня 2017, м. Івано-Франківськ). Івано-Франківський національний технічний університет нафти і газу (сс. 362–364). (Vavruk, E. Ya. (2017). Component-oriented approach in programming java applications. Oil and Gas Energy–2017. Materials of the VI International Scientific and Technical Conference (16 May 2017, Ivano-Frankivsk). Ivano-Frankivsk National Technical University of Oil and Gas (pp. 362–364).).
2. Грищенко, В. М. (2005). Компонентно-орієнтований підхід до розробки програмних систем на основі інтеграції менеджерів інформаційних ресурсів. Проблеми програмування, 1, 3–16. (Grishchenko, V. M. (2005). Component-oriented approach to the development of software systems based on the integration of information resource managers. Problems of programming, 1, 3–16).
3. Кирилецька, Г. (2015). Комп'ютерні технології при обробці результатів педагогічного експерименту. Нова педагогічна думка, 1, 23–25. (Kyryletska, G. (2015). Computer technologies in processing the results of a pedagogical experiment. New pedagogical thought, 1, 23–25).
4. Копей, В. Б. (2017). Компонентно-орієнтоване моделювання кінематики механізмів мовою Python на прикладі механізму верстата-гойдалки. Нафтогазова енергетика–2017. Матеріали VI Міжнародної науково-технічної конференції (16 травня 2017, м. Івано-Франківськ). Івано-Франківський національний технічний університет нафти і газу (сс. 15–19). (Kopey, V. B. (2017). Component-oriented modeling of the kinematics of mechanisms in Python using the example of a swing machine mechanism. Oil and Gas Energy–2017. Materials of the VI International Scientific and Technical Conference (16 May 2017, Ivano-Frankivsk). Ivano-Frankivsk National Technical University of Oil and Gas (pp. 15–19).).
5. Коробейнікова, Т. І., Омельченко, С. С. (2019). Компонентно-орієнтований підхід при розробці ігор. Матеріали XLVIII науково-технічної конференції підрозділів Вінницького національного технічного університету (13-15 березня 2019, м. Вінниця) (сс. 603–605). (Korobeynikova, T. I., Omelchenko, S. S. (2019). Component-oriented approach to game development. Materials of the XLVIII Scientific and Technical Conference of Physicians of the Vinnytsia National Technical University (13-15 February 2019, Vinnytsia) (pp. 603–605).
6. Петрук, В., Семеніхіна, О., Сабодш, Ю. (2022). Нові підходи до статистичного аналізу результатів педагогічного експерименту. Фізико-математична освіта, 33(1), 36–42. (Petruk, V., Semenikhina, O., Sabadosh, Yu. (2022). New approaches to statistical analysis of pedagogical experiment results. Physical and mathematical education, 33(1), 36–42).
7. Ржепецький, В. П. (2011). Статистична обробка результатів експериментальних вимірювань : методичні рекомендації. Кривий Ріг : Криворізький державний педагогічний університет. (Rzhepetsky, V. P. (2011). Statistical processing of experimental measurement results : methodical recommendations. Kryvyi Rih : Kryvyi Rih State Pedagogical University).
8. Рибніков, В. І. (2021). Застосування компонентно-орієнтованого програмування при проектуванні генетичних алгоритмів. Матеріали Першої Всеукраїнської науково-практичної конференції молодих вчених та студентів «Інженерія програмного забезпечення і передові інформаційні технології» (SoftTech-2021) (22-26 листопада 2021, м. Київ). НТУУ «КПІ ім. Ігоря Сікорського». (Rybnikov, V. I. (2021). Application of component-oriented programming in the design of genetic algorithms. Proceedings of the First All-Ukrainian Scientific and Practical Conference of Young Scientists and Students “Software Engineering and Advanced

- Information Technologies” (SoftTech-2021) (November 22-26, 2021, Kyiv). NTUU “Igor Sikorsky Kyiv Polytechnic Institute”).
9. Рубцов, М., Муртазієв, Е., Рубцова, Н. (2018). Методика вибору статистичного критерію та його застосування до результатів педагогічного експерименту. Науковий вісник Мелітопольського державного педагогічного університету, 2(21), 164–72. (Rubtsov, M., Murtaziev, E., Rubtsova, N. (2018). Methodology for choosing a statistical criterion and its application to the results of a pedagogical experiment. Scientific Bulletin of the Melitopol State Pedagogical University, 2(21), 164–72).
  10. Співаковський, О. В. (2004). Теоретико-методичні основи навчання вищої математики майбутніх вчителів математики з використанням інформаційних технологій (автореф. дис. ... докт. пед. наук : 13.00.02). Київ (Spivakovsky, O. V. (2004). Theoretical and methodological foundations of teaching higher mathematics to future mathematics teachers using information technologies (DSc thesis abstract). Kyiv).
  11. Співаковський, О. В., Круглик, В. С. (2005). Технології розробки програмних засобів, які підтримують компонентно-орієнтований підхід. Науковий часопис Українського державного університету імені Михайла Драгоманова. Серія 2. Комп’ютерно-орієнтовані системи навчання, 2(9), 31–42. (Spivakovsky, O. V., Kruglyk, V. S. (2005). Technologies for developing software tools that support a component-oriented approach. Scientific Journal of the Mykhailo Dragomanov Ukrainian State University. Series 2. Computer-based Learning Systems, 2(9), 31–42).
  12. Critical Values of the Mann-Whitney U (Two-Tailed Testing). Retrieved from: <https://ocw.umb.edu/psychology/psych-270/other-materials/RelativeResourceManager.pdf>.
  13. Mann-Whitney U Test. Retrieved from: <https://www.technologynetworks.com/informatics/articles/mann-whitney-u-test-assumptions-and-example-363425>.

**Bazurin V. M., Bazurina S. V. Software implementation of the Mann-Whitney criterion calculation algorithm in C#.**

*The study of methods for statistical processing of experimental results (including pedagogical ones) is carried out. The importance of statistical processing of experimental results is shown. Particular attention is paid to the Mann-Whitney criterion. The algorithm for calculating the Mann-Whitney criterion for two data samples is disclosed. The article also provides a flowchart of the algorithm for calculating the Mann-Whitney criterion for two data samples. The article provides a sequence of developing a program for calculating the Mann-Whitney criterion for two samples. A component-oriented approach and an object-oriented programming paradigm are used. A class diagram of a module that calculates the Mann-Whitney criterion for two samples with an accuracy of  $\alpha=0.05$  is developed. The created module is implemented in the form of a DLL library that is connected to the application being created, and the application can be created in various programming languages that support integration into the DLL library project. The article also describes the fields and methods of the main classes of the created module, provides examples of codes and input data. The developed module can be connected to a software tool that will perform statistical processing of the experiment results.*

**Key words:** *algorithm, software implementation, programming language, Mann-Whitney criterion, software module, DLL library.*