УДК 37.012:519

DOI 10.33251/2522-1477-2021-10-20-26

БОНДАР Ольга Петрівна,

кандидат фізико-математичних наук, доцент,
доцент кафедри фізико-математичних дисциплін,
Львівська академія Національного
авіаційного університету
ORCID 0000-0001-5877-5667

СЕМЕНЮТА Марина Фролівна,

кандидат фізико-математичних наук, доцент,
доцент кафедри фізико-математичних дисциплін,
Львівська академія Національного
авіаційного університету
ORCID 0000-0003-4639-0545

МЕТОДИКА ЗАСТОСУВАННЯ КРИТЕРІЇВ ВІЛКОКСОНА І МАННА-УІТНІ В ПЕДАГОГІЧНОМУ ЕКСПЕРИМЕНТІ

Сформульовано алгоритм застосування критеріїв Вілкоксона і Манна-Уїтні для перевірки гіпотези $H_0: F_1(x) = F_2(x)$ про однорідність вибірок. Розглянуто приклади застосування критеріїв у педагогічному експерименті. Зокрема, обчислено статистичні значення критерію Вілкоксона і критерію Манна-Уїтні, показано їх зв'язок з областю прийняття гіпотези H_0 при різних альтернативних гіпотезах H_1 .

Ключові слова: критерій Вілкоксона, критерій Манна-Уїтні, однорідність вибірок, область прийняття гіпотези, ранг вибірки.

Постановка проблеми. В [1] ми розглянули проблеми ідентифікації і обчислення критеріїв перевірки гіпотези про однорідність двох незалежних вибірок. Зокрема, звернуто увагу на неоднозначні назви критерію перевірки гіпотези, відмінність припущень і формулювань гіпотези, нееквівалентність формул і неповноту висновків в різних джерелах інформації. Також наведено приклад застосування критерію в педагогічних дослідженнях.

Але наш досвід показав, що для набуття стійких навичок практичного застосування, в першу чергу, педагогам-гуманітаріям не достатньо обґрунтованих розрахункових прикладів (в джерелах інформації, що не є суто математичними) застосування зазначених критеріїв.

Відтак, проблема методики застосування критеріїв Вілкоксона і Манна-Уїтні залишається нагальною в науково-педагогічних дослідженнях.

Аналіз останніх досліджень і публікацій. Як показує аналіз Internet-джерел інформації [4; 5], приклади обчислень статистичних значень критеріїв однорідності вибірок в значній мірі носять технічний характер, тобто містять засоби зазвичай комп'ютерної реалізації обчислень. Це є позитивною практикою, але тільки частиною загальної практики використання критеріїв.

Теоретичне підґрунтя, наприклад [2; 3], застосування критеріїв Вілкоксона і Манна-Уїтні, в свою чергу, містить приклади, які не охоплюють всі можливості застосування критеріїв. Відтак, методика використання критеріїв потребує оптимального поєднання теоретичних засад з формуванням практичних навичок перевірки гіпотези про однорідність вибірок.

Мета статті. Нашою метою є надання учасникам освітнього процесу, переважно, здобувачам вищої освіти третього (освітньо-наукового) рівня сформованих на нашому досвіді практичних підходів до розв'язання зазначених проблем. Це стосується, в першу

чергу, проблем методики застосування в науково-педагогічних дослідженнях критеріїв однорідності вибірок.

Виклад основного матеріалу. При проведенні педагогічного експерименту характеристики контрольної та експериментальної груп вважаються репрезентативними вибірками з аналогічних характеристик генеральної сукупності.

Першим етапом статистичної обробки досліджуваних характеристик контрольної та експериментальної груп зазвичай є перевірка цих вибірок на однорідність. Під *однорідністю* розуміють наявність в усіх елементів вибірок таких властивостей, які визначають їх однотипність.

Неоднорідність вибірок може бути зумовлена наявністю похибки оцінки даних або ж присутністю у вибірці «аномальних» елементів. Якщо неоднорідність вибірки зумовлена похибкою вимірювань і дослідник знає її причину, то відповідні елементи вибірки можна виключити перед початком статистичного аналізу даних. В іншому випадку причини неоднорідності потребують досліджень.

Наприклад, при тестуванні двох груп студентів всі студенти отримали «добре» та «відмінно», за винятком одного студента, який отримав «незадовільно». Ці дві вибірки оцінок, вочевидь, не є однорідними. Якщо причиною незадовільної оцінки, наприклад, були суб'єктивні обставини або збій комп'ютерної програми тестування, то є сенс вилучити з вибірки незадовільну оцінку.

Однорідність вибірок не завжди можна визначити за безпосереднім спостереженням їх елементів. Наприклад, дві вибірки оцінок «добре» за шкалою ECTS

Вибірка №1	76	81	81	82	83	86	
Вибірка №2	74	75	76	78	81	82	85

є однорідними, а наступні вибірки оцінок «добре» за тією ж шкалою

Вибірка №1	76	81	83	85	86	88	
Вибірка №2	74	75	76	78	79	80	82

не є однорідними. Безпосереднє порівняння оцінок тут не дає можливості зробити однозначні висновки.

Якщо однорідність вибірок не можна визначити безпосередньо, то її визначають за результатами застосування відповідних статистичних гіпотез.

Нагадаємо, що для перевірки статистичних гіпотез використовується випадкова величина, яку називають статистичним критерієм. Відомі критерії перевірки гіпотези про однорідність вибірок – критерії Вілкоксона і Манна-Уїтні, які іноді об'єднують, називаючи критерієм Вілкоксона-Манна-Уїтні.

Тут важливо, що перевіряються на однорідність незалежні вибірки з генеральної сукупності, що мають неперервні функції розподілу $F_1(x)$ і $F_2(x)$. Однорідність вибірок означає рівність функцій розподілу, відтак основна гіпотеза H_0 про однорідність вибірок полягає в тому, що $F_1(x) = F_2(x)$.

Конкуруючою – альтернативною – гіпотезою при перевірці однорідності вибірок може бути одна з гіпотез: $F_1(x) \neq F_2(x)$, $F_1(x) > F_2(x)$, $F_1(x) < F_2(x)$.

В [1] для кожної з цих гіпотез ми зобразили критичну область, якій належить статистичне значення критерію у разі справедливості альтернативної гіпотези. Звідси випливає визначення області прийняття гіпотези H_0 .

Так, якщо альтернативною є гіпотеза $F_1(x) \neq F_2(x)$, то областю прийняття гіпотези H_0 є відрізок (k_1, k_2) , нижня межа k_1 якого зазвичай називається критичною точкою відповідного критерію і знаходиться з таблиць критичних точок критерію Вілкоксона чи Манна-Уїтні. Верхня межа k_2 обчислюється за формулою.

Якщо альтернативною є гіпотеза $F_1(x) < F_2(x)$, то областю прийняття гіпотези H_0 є промінь $(-\infty, k_{кр})$, тобто множина чисел, менша за критичну точку $k_{кр}$ критерію. Це означає, що якщо статистичне значення критерію виявиться меншим за $k_{кр}$, то гіпотезу про однорідність можна прийняти з певною ймовірністю.

Якщо ж альтернативною є гіпотеза $F_1(x) > F_2(x)$, то областю прийняття гіпотези H_0 є промінь $(k_{кр}, \infty)$, тобто множина чисел, більша за критичну точку $k_{кр}$ критерію. Це означає, що якщо статистичне значення критерію виявиться більшим за $k_{кр}$, то гіпотезу про однорідність можна прийняти з певною ймовірністю. Один з прикладів визначення однорідності вибірок з даною альтернативною гіпотезою подано нижче.

Зауважимо, що якщо кількість елементів хоча б однієї з двох вибірок більша за 25, то для обчислення критеріїв використовується функція Лапласа [2]. Тут же розглядатимемо тільки ситуації, коли кількість елементів двох незалежних вибірок не перевищує 25 і кількість n_1 елементів першої вибірки не більша за кількість n_2 елементів другої вибірки: $n_1 \leq n_2$. Якщо це не так, то вибірки перенумеровуються, тобто міняються місцями. Це потрібно для зручності користування відповідними таблицями.

Алгоритм застосування критеріїв Вілкоксона і Манна-Уїтні для перевірки гіпотези $H_0: F_1(x) = F_2(x)$ про однорідність вибірок.

1. Задати рівень значущості α , який дорівнює ймовірності можливої помилки висновку про однорідність. Відтак, число $1 - \alpha$ визначатиме ймовірність правильного висновку щодо однорідності вибірок.

2. Задати конкуруючу гіпотезу H_1 , за якою визначити область прийняття гіпотези H_0 . Якщо $H_1: F_1(x) \neq F_2(x)$, то областю прийняття гіпотези H_0 є відрізок (k_1, k_2) , де нижню межу k_1 знаходимо з таблиць критичних точок:

критерію Вілкоксона $k_1 = W_{нижнє\ кр}(\alpha/2, n_1, n_2)$,

критерію Манна-Уїтні $k_1 = U_{кр}(\alpha, n_1, n_2)$.

Верхню межу k_2 знаходимо за формулою $k_2 = (n_1 + n_2 + 1) n_1 - k_1$.

Якщо $H_1: F_1(x) > F_2(x)$, то областю прийняття гіпотези H_0 є числа, більші за k_1 , а в разі, якщо $H_1: F_1(x) < F_2(x)$, то областю прийняття гіпотези H_0 є числа, менші за k_1 .

3. Об'єднати обидві вибірки в одну спільну вибірку, розташувавши її елементи у зростаючому або спадному порядку. Пронумерувати елементи об'єднаної вибірки.

4. Визначити для кожного елемента цієї вибірки ранг – номер елемента, якщо елемент не повторюється. Якщо елементи різних вибірок повторюються, то їх рангом є середнє арифметичне їх номерів в об'єднаній вибірці.

5. Знайти суму R_1 рангів елементів першої вибірки.

6. Визначити статистичне значення критерію.

Статистичне значення W критерію Вілкоксона – це сума рангів елементів першої вибірки: $W = R_1$.

Статистичне значення U критерію Манна-Уїтні знаходимо наступним чином:

– обчислюємо суму R_2 рангів елементів другої вибірки;

– визначаємо найбільший з рангів $R = \max(R_1, R_2)$ і кількість n елементів вибірки з найбільшим рангом;

– обчислюємо значення U критерію Манна-Уїтні за формулою

$$U = n_1 \cdot n_2 + \frac{n \cdot (n + 1)}{2} - R.$$

7. Сформулювати висновок. Якщо знайдене статистичне значення W або U відповідного критерію належить області прийняття гіпотези, то з ймовірністю $1 - \alpha$ можна стверджувати, що вибірки однорідні, в іншому разі вибірки не є однорідними.

Визначений нами алгоритм застосування критеріїв продемонструємо на згаданих прикладах вибірок оцінок «добре» за шкалою ECTS.

Приклад 1. Планується провести педагогічний експеримент в групі з 7 студентів і порівняти його результат з показниками контрольної групи, яка складається з 6 студентів. На початку експерименту оцінки x_i студентів контрольної групи і оцінки y_j студентів експериментальної групи є такими:

x_i	76	81	81	82	83	86	
y_j	74	75	76	78	81	82	85

Для об'єктивності дослідження на початку експерименту групи мають бути однорідними. Відтак, за рівнем значущості $\alpha = 0,05$ перевіримо гіпотезу $H_0: F_1(x) = F_2(x)$ про однорідність вибірок оцінок x_i і y_j . За конкуруючу гіпотезу H_1 приймемо $F_1(x) \neq F_2(x)$ при перевірці за критерієм Вілкоксона і гіпотезу $F_1(x) > F_2(x)$ при перевірці за критерієм Манна-Уїтні.

Перевірка за критерієм Вілкоксона. Оскільки у вибірці оцінок студентів контрольної групи 6 оцінок, а у вибірці оцінок студентів експериментальної групи більше оцінок – 7, то першою вважатимемо вибірку з меншою кількістю елементів, тобто $n_1=6$, тоді $n_2=7$. Зауважимо, що якщо б вибірки мали однакову кількість елементів, то першою можна було б вважати будь-яку з вибірок.

За умовою конкуруючою гіпотезою є $H_1: F_1(x) \neq F_2(x)$, тому областю прийняття гіпотези H_0 буде інтервал. Нижню його межу знайдемо з таблиці критичних точок критерію Вілкоксона:

$$k_1 = W_{\text{нижнє кр.}}(0,05/2, 6, 7) = W_{\text{нижнє кр.}}(0,025, 6, 7) = 27.$$

Тоді верхня межа дорівнює: $k_2 = (n_1 + n_2 + 1)n_1 - k_1 = (6 + 7 + 1)6 - 27 = 84 - 27 = 57$.

Відтак, областю прийняття гіпотези H_0 є відрізок (27, 57).

Обчислимо статистичне значення W критерію Вілкоксона, для чого об'єднаємо обидві вибірки, записавши їх елементи у зростаючому порядку:

Об'єдн. вибірка 74 75 76 76 78 81 81 81 82 82 83 85 86

Пронумеруємо елементи цієї вибірки за їх порядком і виділимо елементи першої вибірки:

<i>Об'єдн. вибірка</i>	74	75	76	76	78	81	81	81	82	82	83	85	86
<i>Пор. номер</i>	1	2	3	4	5	6	7	8	9	10	11	12	13

Кожному елементу, що не повторюється, поставимо у відповідність ранг – порядковий номер елемента. А елемент 76 **р і з н и х** вибірок повторюється, тому його рангом буде середнє арифметичне його порядкових номерів 3 і 4:

$$\frac{3+4}{2} = 3,5.$$

Аналогічно визначається ранг елемента 81 $\frac{7+8}{2} = \frac{15}{2} = 7,5$ і ранг елемента 82 є таким: $\frac{9+10}{2} = 9,5$.

Запишемо ранги елементів, виділивши серед них ранги елементів першої вибірки

<i>Об'єдн. виб.</i>	74	75	76	76	78	81	81	81	82	82	83	85	86
<i>Пор. номер</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>Ранг ел-та</i>	1	2	3,5	3,5	5	6	7,5	7,5	9,5	9,5	11	12	13

Знайдемо суму рангів елементів першої вибірки:

$$R_1 = 3,5 + 6 + 7,5 + 9,5 + 11 + 13 = 50,5.$$

Таким чином, обчислено статистичне значення W критерію Вілкоксона

$$W = R_1 = 50,5.$$

Оскільки воно належить області $(27, 57)$ прийняття гіпотези, тобто

$$50,5 > 27 \text{ і } 50,5 < 57,$$

то з ймовірністю $1 - 0,05 = 0,95$ можна стверджувати, що вибірки однорідні.

Перевірка за критерієм Манна-Уїтні. Оскільки за конкуруючу гіпотезу H_1 прийнято $F_1(x) > F_2(x)$, то областю прийняття гіпотези H_0 є числа, більші за критичну точку k_1 . Знайдемо її за таблицями критичних точок критерію Манна-Уїтні: $k_1 = U_{\text{кр}}(\alpha, n_1, n_2) = U_{\text{кр}}(0,05; 6, 7) = 6$.

Відтак, областю прийняття гіпотези H_0 є числа, більші за $k_1 = 6$.

Знайдемо суму рангів елементів другої вибірки:

$$R_2 = 1 + 2 + 3,5 + 5 + 6,5 + 9,5 + 12 = 39,5.$$

Більший з рангів $R = \max(R_1, R_2) = \max(51,5; 39,5) = 51,5$ є рангом першої вибірки, тому n дорівнює кількості її елементів $n = 6$.

Обчислюємо статистичне значення U критерію Манна-Уїтні:

$$U = n_1 \cdot n_2 + \frac{n \cdot (n+1)}{2} - R = 6 \cdot 7 + \frac{6 \cdot 7}{2} - 51,5 = 42 + 21 - 51,5 = 11,5.$$

Оскільки воно належить області $U > U_{\text{кр}}$ прийняття гіпотези, тобто $11,5 > 6$, то з ймовірністю $1 - 0,05 = 0,95$ можна стверджувати, що вибірки однорідні.

Ми бачимо, що висновки про однорідність вибірок в прикладі 1 за критерієм Вілкоксона і критерієм Манна-Уїтні збігаються.

Приклад 2. Нехай на початку експерименту оцінки x_i студентів контрольної групи і оцінки y_j студентів експериментальної групи є такими:

Вибірка №1	76	81	83	85	86	88	
Вибірка №2	74	75	76	78	79	80	82

Так, як і в прикладі 1, за рівнем значущості $\alpha = 0,05$ перевіримо гіпотезу $H_0: F_1(x) = F_2(x)$ про однорідність вибірок оцінок x_i і y_i . За конкуруючу гіпотезу H_1 приймемо $F_1(x) \neq F_2(x)$ при перевірці за критерієм Вілкоксона і гіпотезу $F_1(x) > F_2(x)$ при перевірці за критерієм Манна-Уїтні.

Перевірка за критерієм Вілкоксона. Аналогічно прикладу 1, отримаємо:

Об'єдн. виб.	74	75	76	76	78	79	80	81	82	83	85	86	88
Пор. номер	1	2	3	4	5	6	7	8	9	10	11	12	13
Ранг ел-та	1	2	3,5	3,5	5	6	7	8	9	10	11	12	13

Статистичне значення W критерію Вілкоксона

$$W = R_1 = 3,5 + 8 + 10 + 11 + 12 + 13 = 57,5$$

не належить області $(27, 57)$ прийняття гіпотези тому, що $57,5 > 57$.

Відтак, з ймовірністю $1 - 0,05 = 0,95$ можна стверджувати, що вибірки не є однорідними.

Перевірка за критерієм Манна-Уїтні. Аналогічно прикладу 1, отримаємо:

$$R_2 = 1 + 2 + 3,5 + 5 + 6 + 7 + 9 = 33,5.$$

$$R = \max(R_1, R_2) = \max(57,5; 33,5) = 57,5$$

є рангом першої вибірки, тому n дорівнює кількості її елементів $n = 6$.

Статистичне значення U критерію Манна-Уїтні:

$$U = 6 \cdot 7 + \frac{6 \cdot 7}{2} - 57,5 = 42 + 21 - 57,5 = 5,5.$$

З таблиці критичних точок критерію Манна-Уїтні знаходимо

$$U_{кр} = U(0,05, 6, 7) = 6.$$

Оскільки $U < U_{кр}$, тобто $5,5 < 6$, то з ймовірністю $1 - 0,05 = 0,95$ можна стверджувати, що вибірки не однорідні. Висновки про неоднорідність вибірок в прикладі 2 за критерієм Вілкоксона і критерієм Манна-Уїтні також збігаються.

Протилежні висновки щодо однорідності вибірок у прикладах 1 і 2 свідчать про суттєву відмінність цих вибірок. Однорідність вибірок у першому прикладі означає, що відмінності елементів вибірок можна вважати не статистично достовірними, тобто вони мають випадковий характер. Неоднорідність вибірок у другому прикладі означає, що відмінності їх елементів можна вважати статистично стійкими. Іншими словами, перетин інтервалів обох вибірок в першому прикладі, тобто зона збігу оцінок, є значно більшою за зону збігу у другому прикладі. Так, якщо у другому прикладі з першої вибірки вилючити 76, а з другої – 82, то всі оцінки студентів першої групи будуть вищими за оцінки студентів другої групи, що вже наочно показує неоднорідність вибірок.

Висновки та перспективи подальших досліджень. Нами сформульовано алгоритм застосування критеріїв Вілкоксона і Манна-Уїтні для перевірки гіпотези $H_0: F_1(x) = F_2(x)$ про однорідність вибірок. Розглянуто приклади застосування критеріїв у педагогічному експерименті. Зокрема, обчислено статистичні значення критерію W Вілкоксона і критерію U Манна-Уїтні, показано їх зв'язок з областю прийняття гіпотези H_0 при різних альтернативних гіпотезах H_1 .

Перспективою наших подальших досліджень є формування методики застосування критеріїв найбільш вживаних в науково-педагогічних дослідженнях розподілів випадкових величин.

Список використаних джерел

1. Бондар О. П., Якуніна І. Л., Задорожна О. В. Критерії Вілкоксона і Манна-Уїтні в педагогічних дослідженнях. Науковий вісник Льотної академії. Серія: Педагогічні науки. Збірник наукових праць. Кропивницький. 2018. Вип. 4. С. 53–57.
2. Гмурман В. Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов. Москва: Высш. школа, 2003. 479с.
3. Новиков Д. А. Статистические методы в педагогических исследованиях (типовые случаи). Москва: МЗ-Пресс, 2004. 67с.
4. Онлайн расчет критерия U-Манна-Уитни. URL: <https://statpsy.ru/mana-uitni/online-raschet-kriteriya-u-manna-uitni/> (дата звернення 02.09.2021).
5. Пример расчета критерия U-Манна-Уитни. URL: <https://statpsy.ru/mana-uitni/primer-manna-uitni/> (дата звернення 02.09.2021).

References

1. Bondar, O.P., Yakunina, I.L., Zadorozhna, O.V. (2018). Kryteriyi Vilkoksona i Manna-Uitni v pedahohichnykh doslidzhennyakh [Wilcoxon and Mann-Whitney criteria in pedagogical research]. Naukovyy visnyk L'otnoyi akademiyi. Seriya: Pedahohichni nauky. Zbirnyk naukovykh prats'. Kropyvnyts'kyu. Vyp. 4. S. 53-57 [in Ukrainian].
2. Gmurman, V.E. (2003). Teoriya veroyatnostey i matematicheskaya statistika: ucheb. posobiye dlya vuzov [Probability theory and mathematical statistics: textbook. manual for universities]. Moskva: Vyssh. shkola. 479 s. [in Russian].
3. Novikov, D.A. (2004). Statisticheskiye metody v pedagogicheskikh issledovaniyakh (tipovyye sluchai) [Statistical methods in pedagogical research (typical cases)]. Moskva: M3-Press. 67 s. [in Russian].
4. Onlayn raschet kriteriya U-Manna-Uitni. [Online calculation of the U-Mann-Whitney test]. Retrieved from: <https://statpsy.ru/mana-uitni/online-raschet-kriteriya-u-manna-uitni/> (Last accessed: 02.09.2021) [in Russian].

5. Primer rascheta kriteriya U-Manna-Uitni [Example of calculation of the U-Mann-Whitney criterion]. Retrieved from: <https://statpsy.ru/mana-uitni/primer-manna-uitni/> (Last accessed: 02.09.2021) [in Russian].

BONDAR Olga, The Candidate of Physics and Mathematics, Associate Professor, Associate Professor at the Department of Physical and Mathematical Disciplines, Flight Academy of National Aviation University;

SEMENYUTA Marina, The Candidate of Physics and Mathematics, Associate Professor, Head of the Department of Physical and Mathematical Disciplines, Flight Academy of National Aviation University.

METHODS OF APPLICATION OF WILCOXONE AND MANNA-WHITNEY CRITERIA IN PEDAGOGICAL EXPERIMENT

Abstract. In [1], we considered the problems of identifying and calculating the criteria for testing the hypothesis of homogeneity of two independent samples. In particular, we drew attention to the ambiguous names of the hypothesis test criterion. We also noted the difference between the assumptions and formulations of the hypothesis. In this case, different formulas for calculations require proof of their equivalence.

Our analysis of research and publications has shown that examples of the application of the criteria of homogeneity of Wilcoxon and Mann-Whitney samples are largely technical in nature. That is, these examples usually contain tools for calculating statistics and do not cover all the possibilities of applying the criteria.

Therefore, the method of using criteria requires an optimal combination of theoretical principles with the formation of practical skills of hypothesis testing. When conducting a pedagogical experiment, the researcher should be able to consider theoretically sound calculation examples. These examples of the application of criteria should be available primarily to researchers who are not mathematicians.

Our experience has confirmed the need for the researcher to acquire stable skills in the practical application of the criteria of sample homogeneity in scientific and pedagogical research.

We considered the concept of sample homogeneity in conducting a pedagogical experiment on the examples of statistical information of control and experimental groups. It was noted that the homogeneity of the samples can not always be determined by direct observation of their elements.

We formulated an algorithm for applying the Wilcoxon and Mann-Whitney criteria to test the hypothesis $H_0: F_1(x) = F_2(x)$ about the homogeneity of the samples. We have considered examples of the application of criteria in a pedagogical experiment. In particular, the statistical values of the Wilcoxon test and the Mann-Whitney criteria were calculated. Also, we show their connection with the area of acceptance of hypothesis H_0 under different alternative hypotheses H_1 .

The prospect of our further research is the formation of methods for applying the criteria most commonly used in scientific and pedagogical research of distributions of random variables.

Key words: Wilcoxon criteria, Mann-Whitney criteria, sample homogeneity, hypothesis acceptance area, sample rank.

Одержано редакцією: 20.08.2021 р.
Прийнято до публікації: 27.08.2021 р.