

**Філоненко Н.В.,**  
кандидат фіз.-мат. наук, доцент кафедри математики, інформатики і методики  
навчання  
Переяслав-Хмельницький державний педагогічний  
університет імені Григорія Сковороди

## **КОМП'ЮТЕРНА ОБРОБКА СТАТИСТИЧНОГО КРИТЕРІЮ МАННА – УІТНІ**

Одним із статистичних критеріїв, які вивчаються в курсі математичної статистики для практичних психологів є непараметричний критерій Манна – Уітні, призначений для оцінювання відмінностей між двома вибірками за рівнем якої-небудь ознаки, кількісно вимірної. Даний критерій є непараметричним аналогом t-критерію Стюдента для незалежних вибірок, які не розподілені за нормальним законом. Він дозволяє виявити відмінності між малими вибірками і є більш потужним, ніж інші критерії. В кожній вибірці повинно бути не менше трьох значень.

Як правило, першою вибіркою вважається ряд, в якому значення, за попередньою оцінкою, вищі, ніж значення другого ряду. При цьому висувуються дві гіпотези:

**$H_0$ :** рівень ознаки в ряді 2 не нижчий рівня ознаки в ряді 1 (тобто вибірки належать одній генеральній сукупності).

**$H_1$ :** рівень ознаки в ряді 2 нижчий рівня ознаки в ряді 1.

Ідея критерію Манна – Уітні полягає у представленні всіх значень двох вибірок у вигляді однієї упорядкованої послідовності значень, ранжованої від найменшого до найбільшого значення. При цьому ранги значень, що співпадають, обчислюються як середнє арифметичне порядкових номерів даних значень.

Обчислення при невеликих об'ємах вибірок нескладні, але вже при  **$n_1, n_2 > 20$**  ранжування стає достатньо трудомістким завданням. У цьому

випадку доцільно використовувати комп'ютерну обробку для обчислення значення критерію  $U$ . Можна скористатись програмами комп'ютерної статистики SPSS, Statistica [2] або проводити обчислення з використанням електронних таблиць Excel [1]. Відмітимо, що використання електронних таблиць, на нашу думку, є більш доцільним, оскільки не вимагає наявності спеціальних прикладних програм на комп'ютерах студентів і дозволяє поглибити знання з використання ряду функцій, закладених в електронних таблицях.

Порядок обчислення критерію Манна – Уїтні при використанні електронних таблиць дещо відрізняється від ручного [4] і є наступний.

1. Об'єднуємо обидві вибірки в одну. Об'єм об'єднаної вибірки буде  $n = n_1 + n_2$ . Ранжуємо об'єднану вибірку, приписуючи меншому значенню менший ранг.

2. Знаходимо ранги  $R_i$  об'єднаної вибірки.

3. Знаходимо окремо суми рангів першої і другої вибірок:

$$R_1 = \sum R_{1i}, \quad R_2 = \sum R_{2i}$$

4. Визначаємо розрахункові значення критерію за формулами:

$$U_1 = n_1 n_2 + 0,5 n_1 (n_1 + 1) - R_1$$

$$U_2 = n_1 n_2 + 0,5 n_2 (n_2 + 1) - R_2$$

де  $n_1$  - об'єм першої вибірки,

$n_2$  – об'єм другої вибірки,

Для перевірки правильності розрахунку використовуємо співвідношення

$$U_1 + U_2 = n_1 n_2.$$

5. Знаходимо значення критерію за формулою  $U_{\min} = \min(U_1, U_2)$ .

6. Визначаємо критичне значення  $U_{кр}$  за таблицею для заданого рівня значущості. Якщо  $U_{\text{емп}} > U_{кр}$ , приймається нульова гіпотеза. В протилежному випадку приймається альтернативна гіпотеза про існування статистично значущої різниці між рівнями ознаки у двох вибірках.

Розглянемо використання електронних таблиць Excel для обчислення значення критерію Манна – Уїтні для даних наступного прикладу.

#### Приклад.

Використовуючи критерій Манна – Уїтні, перевірити існування статистично значущої різниці між рівнями ознаки у двох вибірках (табл. 1) на рівні значущості 0,05.

Таблиця 1

Вибірка 1	75	30	25	10	30	20	50
Вибірка 2	10	10	15	20	30	25	5

У стовпцях А та В вводимо відповідно вибірки 1 та 2.

У стовпці С формуємо об'єднану вибірку. У стовпці D обчислюємо ранги для об'єднаної вибірки. Для цього в комірку D3 поміщаємо функцію =РАНГ.СП(C3;\$C\$3:\$C\$16;1), після чого розмножуємо її для всього стовпця.

Далі виконуємо розрахунки всіх необхідних значень за наведеним алгоритмом. Для розрахунків необхідні функції СУММ – для підрахунку суми рангів для першої та другої вибірок, ЧСТРОК – для підрахунку об'ємів вибірок, НАИМЕНШИЙ – для знаходження меншого із значень  $U_1, U_2$ , ЕСЛИ – для порівняння експериментального та критичного значень критерію.

	A	B	C	D	E	F	G
1	значення ознаки	об'єднана ранги для					
2	вибірка 1	вибірка 2	вибірка	об'єдн. вибірки			
3	75	10	75	14			
4	30	10	30	11	R1=	67	
5	25	15	25	8,5	R2=	38	
6	10	20	10	3			
7	30	30	30	11	U1=	10	
8	20	25	20	6,5	U2=	39	
9	50	5	50	13	Uemp=	10	
10			10	3	U1+U2=	49	
11			10	3	n1*n2=	49	
12			15	5	Uкр=	11	
13			20	6,5	висновок	H1	
14			30	11			
15			25	8,5			
16			5	1			

В результаті обчислень робимо висновок про те, що нульова гіпотеза відхиляється і приймається альтернативна гіпотеза  $H_1$  про існування статистично значущої різниці між рівнями ознаки у двох вибірках.

### Список використаної літератури

1. Лапач С.Я., Губенко А.В., Бабиш П.Н. Статистические методы в медико- биологических исследованиях с использованием Excel - К: МОРИОН. 2000. -320 с.
2. Наследов А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. Учебное пособие. 2-е изд., - СПб.: Речь, 2006. - 392 с.
3. Сидоренко Е.В. Методы математической обработки в психологии. –Санкт-Петербург: Социально- психологический центр, 1996.-350с.
4. Філоненко Н.В. Василькевич Я.З. Математична статистика: науково-практичний посібник – Переяслав-Хмельницький, 2002. – 48 с.