

УДК 796.011

ОСОБЕННОСТИ ПРИМЕНЕНИЯ КРИТЕРИЯ МАННА – УИТНИ ДЛЯ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ В ФИЗИЧЕСКОЙ КУЛЬТУРЕ И СПОРТЕ

Абдрахманова И.В., Лущик И.В.

ФГБОУ ВО «Волгоградская государственная академия физической культуры», Волгоград,
e-mail: academy@vgafk.ru

В статье описаны проблемы проверки статистических гипотез об отсутствии значимых различий между выборочными данными малого объема. Современные диссертационные исследования предполагают многоаспектное сравнение результатов испытаний. При этом обрабатываются данные, полученные в контрольной и экспериментальной группах, а также в группах, отличающихся по половому, возрастному или другим признакам. В области физической культуры и спорта в качестве признака группировки может рассматриваться уровень технико-тактической подготовки или спортивного мастерства, квалификация спортсмена. Данные условия обуславливают малый объем числовых массивов, подлежащих статистической обработке. Авторами рассмотрены непараметрические критерии, наиболее часто применяемые молодыми исследователями для обоснования эффективности экспериментальных воздействий. Описаны принципы построения соответствующих математических моделей и варианты интерпретаций прогнозируемых результатов. В статье представлены результаты статистического анализа физических качеств детей и подростков различных возрастных групп и рекомендации по применению критерия Манна – Уитни. Выявлена неоднозначность подходов к формализации данного критерия в современных научных трудах. Сопоставлены выводы, формулируемые при использовании различных моделей его расчета для идентичных массивов значений. Определены причины, обуславливающие противоречивость результатов проверки статистических гипотез.

Ключевые слова: статистический анализ, физическая культура, непараметрические критерии, проверка статистических гипотез

FEATURES OF THE APPLICATION OF THE MANN – WHITNEY CRITERION FOR TESTING STATISTICAL HYPOTHESES IN PHYSICAL CULTURE AND SPORTS

Abdrakhmanova I.V., Luschik I.V.

Volgograd State Physical Education Academy, Volgograd, e-mail: academy@vgafk.ru

The article describes the problems of testing statistical hypotheses about the absence of significant differences between small sample data. Modern dissertation research involves a multidimensional comparison of test results. At the same time, the data obtained in the control and experimental groups, as well as in groups differing in gender, age or other characteristics, are processed. In the field of physical culture and sports, the level of technical and tactical training or sportsmanship, the qualification of an athlete can be considered as a sign of grouping. These conditions cause a small amount of numeric arrays to be statistically processed. The authors consider nonparametric criteria most often used by young researchers to justify the effectiveness of experimental effects. The principles of constructing appropriate mathematical models and options for interpreting the predicted results are described. The article presents the results of a statistical analysis of the physical qualities of children and adolescents of various age groups and recommendations for the application of the Mann-Whitney criterion. The ambiguity of approaches to the formalization of this criterion in modern scientific works is revealed. The conclusions formulated when using different models of its calculation for identical arrays of values are compared. The reasons for the inconsistency of the results of testing statistical hypotheses are determined.

Keywords: statistical analysis, physical education, nonparametric criteria, statistical hypothesis testing

Современные труды молодых ученых в области физической культуры и спорта разнообразны, но чаще всего посвящены проблемам совершенствования процесса формирования физических качеств у представителей различных возрастных групп или организации тренировочной деятельности на основе применения инновационных подходов. Эффективность разрабатываемых методических рекомендаций или построенных моделей может быть подтверждена посредством проверки статистических гипотез. Наиболее распространенным ин-

струментом в этом случае является критерий Стьюдента, имеющий две модели:

– для сравнения данных в случае связанных выборок (например, сопоставление данных по одной и той же группе, зарегистрированных до и после экспериментального воздействия);

– для несвязанных выборок (в частности, выявление значимых различий в контрольной и экспериментальной группах).

Исследователи объясняют востребованность данного критерия его многофункциональностью: он позволяет оценить

однородность выборок (доказать их принадлежность к одной генеральной совокупности), значимость различий между средними выборочными. Применение критерия позволяет построить интервальную оценку генеральной средней или средней дисперсии на основе взаимосвязи трех основных характеристик выборочной совокупности: «ширины доверительного интервала, соответствующей ему доверительной вероятности и объема выборки» [1, с. 43].

Перечисленные преимущества t-критерия Стьюдента в области физической культуры и спорта не имеют такой выраженности, так как его использование ограничено двумя требованиями:

- нормальное распределение сопоставляемых выборочных данных;
- равенство дисперсий.

В научно-методической литературе описываются трудности проверки нормальности распределения и равенства дисперсий, а также последствия нарушений хотя бы одного из этих условий. В частности, одним из условий корректности проверки является существенное увеличение объема выборок (до 2500), что невыполнимо для малых массивов данных. В этих условиях рекомендуется использование критерия Крамера – Уэлча, где «разность выборочных средних арифметических для двух выборок делится на естественную оценку среднего квадратического отклонения этой разности» [2, с. 12].

Если сопоставление производится для большего количества групп или объем выборок относительно мал, то решение данной задачи при помощи параметрических критериев в ряде случаев некорректно [3, с. 120].

Отбор непараметрических критериев для сопоставления результатов измерений осуществляется на основе учета следующих факторов: количество сопоставляемых выборок и объем каждой из сопоставляемых выборок. При этом необходимо также учитывать относительную мощность отбираемых критериев. Если сопоставляются две выборки с целью выявления различий между ними по уровню одного исследуемого признака, то чаще всего используются два непараметрических критерия. Молодые исследователи испытывают сложности при выборе между критерием Розенбаума и критерием Манна – Уитни.

С целью решения представленных выше проблем был произведен сравнительный анализ непараметрических критериев, используемых в современной практике статистического анализа данных малого объема.

Цель исследования – определить условия, при которых использование критерия Манна – Уитни будет целесообразным; опи-

сать приведенные в специальной литературе подходы к расчету значения критерия Манна – Уитни и выявить возможные несоответствия между результатами их применения.

Материалы и методы исследования

При проведении исследования осуществлен сравнительный анализ материалов, представленных в открытых информационных источниках. Описанные математические модели применены для обработки данных по оценке уровня физического развития подростков. Произведено сопоставление результатов в формальном и графическом представлении при помощи онлайн-калькуляторов.

Результаты исследования и их обсуждение

Проблема выбора критерия для сравнения данных выборок малого объема по уровню исследуемого признака широко рассматривается в современной специальной литературе. Для параметрических критериев необходима первоначальная проверка нормального распределения сопоставляемых данных [4, с. 56].

Данная процедура связана с использованием целого ряда критериев, имеющих собственную квалификацию и систему ограничений применимости [5, с. 83].

Указанные трудности игнорируются при проверке гипотезы о равенстве выборочных средних посредством использования непараметрических критериев.

Если сопоставляются данные для двух зависимых выборок малого объема, то чаще всего используется Q-критерий Розенбаума. Такой отбор определяется условием сопоставимости объема сравниваемых выборок (в частности, их равенства). Существенным недостатком Q-критерия является ограничение объема выборок (не более 10 наблюдений в каждом).

U-критерий Манна – Уитни является более мощным. Достаточно, чтобы объем каждой выборки превысил два наблюдения при верхней границе 60 наблюдений, определяемой таблицей критических значений. Возможно применение U-критерия даже при условии, когда объем одной выборки равен 2, а объем другой больше 4.

Позиционирование критерия Манна – Уитни в системе отбора непараметрических критериев представлено в научно-методической литературе (рис. 1).

Для использования критерия необходимо объединить выборочные данные и проранжировать их. Затем следует вновь разделить выборки. Критерий сводится к оценке зоны пересечения проранжированных данных. Чем эта зона меньше, тем существеннее различия.

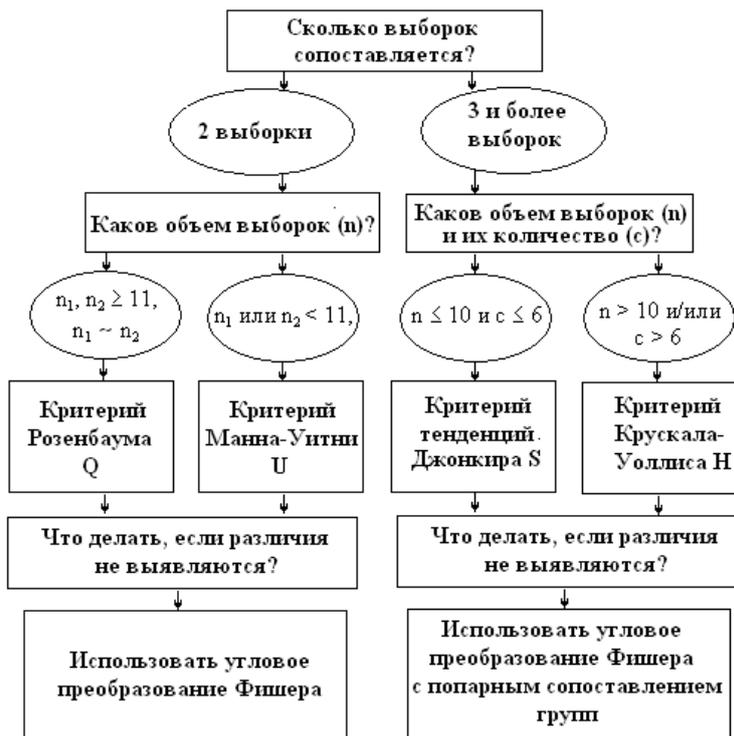


Рис. 1. Система отбора непараметрического критерия для определения различия в выборках по уровню исследуемого признака

Расчетное значение находится следующим образом:

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

где n_1 – объем первой выборки;
 n_2 – объем второй выборки;
 n – объем выборки, имеющей большую ранговую сумму;
 T – большая сумма рангов выборок [6, с. 52]

В открытых информационных источниках представлен еще один вариант определения расчетного значения U-критерия:

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

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

$$U = \min\{U_1; U_2\},$$

где R_1 – ранговая сумма первой выборки;
 R_2 – ранговая сумма второй выборки [7].

Вопрос идентичности результатов использования расчетных формул не рассмотрен подробно в современных научных трудах.

Приведем пример расчета U-критерия для выборочных данных, не имеющих нор-

мального распределения. При исследовании реакции выбора у мальчиков-подростков был произведен подсчет количества ошибок, совершаемых в стандартных условиях. В первой группе количество тренировок превышало 13 ч в неделю, во второй объем тренировок был меньше 13 ч в неделю. Результаты расчетов представлены в таблице.

I способ расчета:

$$U_1 = 8 \cdot 12 + \frac{8 \cdot 9}{2} - 88 = 44,$$

$$U_2 = 8 \cdot 12 + \frac{12 \cdot 12}{2} - 122 = 52,$$

$$U = \min\{44; 52\} = 44.$$

II способ расчета:

$$U = 8 \cdot 12 + \frac{12 \cdot 12}{2} - 122 = 52.$$

Наблюдаемые различия в результатах расчетов могут быть проигнорированы, так как в обоих случаях происходит отклонение нулевой гипотезы (критическое значение U-критерия равно 22, то есть существенно меньше расчетного). Однако отсутствие идентичности результатов является значительным недостатком критерия.

N	Значения в первой группе	N	Значения во второй группе
1	0	1	0
2	0	2	0
3	0	3	0
4	0	4	0
5	0	5	0
6	1	6	1
7	0	7	0
8	1	8	0
		9	0
		10	0
		11	1
		12	0

Рассчитать показатели

U-критерий Манна-Уитни равен 52
 Критическое значение U-критерия Манна-Уитни при заданной численности сравниваемых групп составляет 22
 52 > 22, следовательно различия уровня признака в сравниваемых группах статистически не значимы (p>0.05)

Рис. 2. Результаты расчета U-критерия при помощи программы medstatistic.ru/calculators

Современные сервисы обеспечивают возможность онлайн-расчета значения критерия, при этом отмечается, что результаты могут существенно отличаться в зависимости от того, учтена ли поправка на непрерывность. В случае аппроксимации генеральной совокупности и при рассмотрении выборки малого объема с дискретными значениями могут быть получены заключения о значимости и незначимости различий связанных рангов для одних и тех же массивов данных.

В рассмотренном случае очевидно, что различия несущественны, так как наблюдается тотальное пересечение рядов рангов. При этом результаты калькуляции, представленные бесплатными программами расчета, отличаются (рис. 2, 3).

Описанные выше несоответствия расчетных характеристик процедуры проверки статистической гипотезы об отсутствии значимых различий средних значений для независимых выборок на основе использования непараметрического критерия Манна – Уитни предполагают многоаспектный подход к анализу экспериментальных данных. В частности, необходима более глубокая детализация в случае использования электронных систем оценки. Для дискретных характеристик следует применять

способ расчета U-критерия для каждой из ранговых сумм и минимизации полученных результатов.

Result Details

Sample 1
 Sum of ranks: 88
 Mean of ranks: 11
 Expected sum of ranks: 84
 Expected mean of ranks: 10.5
 U-value: 44
 Expected U-value: 48

Sample 2
 Sum of ranks: 122
 Mean of ranks: 10.17
 Expected sum of ranks: 126
 Expected mean of ranks: 10.5
 U-value: 52
 Expected U-value: 48

Sample 1 & 2 Combined
 Sum of ranks: 210
 Mean of ranks: 10.5
 Standard Deviation: 12.9615

Рис. 3. Результаты расчета U-критерия при помощи программы Mann – Whitney U Test Calculator

Заключение

«Ручной» способ ранжирования данных с количеством градаций, превышающим 20, и их дальнейшей обработки является трудоемким и нецелесообразным. Ранее в этом случае для сравнения выборочных средних применялось угловое преобразование Фишера. Современные сервисы предоставляют возможность расчета критерия Манна – Уитни для малых выборок, не имеющих нормального распределения с относительно большим количеством градаций. Данная процедура предполагает аппроксимацию данных к непрерывному распределению, что может дать некорректный или неоднозначный результат. Для исключения таких трудностей следует использовать сопоставление результатов использования двух схем расчета. Возможно представление ранжированных значений посредством применения возможностей графических редакторов. Визуальная оценка ширины зоны перекрытия ранговых интервалов в совокупности с расчетными значениями обеспечивает повышение достоверности выводов и снижает

вероятность ошибок при проверке статистических гипотез.

Список литературы

1. Шаталов К.В., Кириллова А.В. Применение критерия Стьюдента для оценки результатов межлабораторных сравнительных испытаний // Эталоны. Стандартные образцы. 2016. № 1. С. 42-48.
2. Орлов А.И. Проверка статистической гипотезы однородности математических ожиданий двух независимых выборок: критерий Крамера – Уэлча вместо критерия Стьюдента // Научный журнал КубГАУ. 2015. № 110. С. 1–22.
3. Абдрахманова И.В., Лущик И.В. Проблемы проверки статистических гипотез при исследовании малых выборок в физической культуре и спорте // Современные наукоемкие технологии. 2021. № 12–1. С. 119–123.
4. Унгурияну Т.Н., Гржибовский А.М. Краткие рекомендации по описанию, статистическому анализу и представлению данных в научных публикациях // Экология человека. 2011. № 5. С. 55–60.
5. Александровская Л.Н., Кириллин А.В. Рекомендации по применению ряда критериев проверки отклонения распределения вероятностей от нормального закона в практике инженерного статистического анализа // Известия Самарского научного центра РАН. 2017. № 1–1. С. 82–90.
6. Сидоренко Е.В. Методы математической обработки в психологии. СПб.: ООО «Речь», 2002. 350 с.
7. U-критерий Манна – Уитни // Википедия. [2022]. Дата обновления: 12.03.2022. URL: <https://ru.wikipedia.org/?curid=620166&oldid=120592421> (дата обращения: 03.04.2022).