РОЗДІЛ 1

СУЧАСНИЙ РОЗВИТОК І ВИКЛАДАННЯ БІОМЕХАНІКИ ФІЗИЧНОГО ВИХОВАННЯ ТА СПОРТУ

УДК 796.01

Волков Ю. О., Екимов В. Ю., Шиндер М. В.

КРИТЕРИИ ОЦЕНКИ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ В БИОМЕХАНИКЕ

В практике спортивных исследований статистические методы позволяют решить следующую типовую задачу: две группы спортсменов, занимающиеся одним и тем же видом спорта, тренировались с использованием различных (традиционной и экспериментальной) методик. Для оценки сравнительной эффективности экспериментальной методики необходимо выявить отсутствие или наличие статистических различий между группами по какому-либо признаку. Задача решается с помощью стандартных статистических методов, но с учётом того, что сферой приложения является область спорта.

Ключевые слова: *статистика, гипотезы, критерии, экспериментальные исследования, биомеханика.*

Постановка проблемы. Анализ последних исследований и публикаций. Крупный австралийский математик, специалист в области математической статистики, профессор Э. Питмен пришел к заключению, что в теории статистических выводов необходимо учитывать два момента. «*Во-первых*, …сформулировать принципы, которые помогут статистикам *оценить качество данных* эксперимента или испытания в подтверждение (или опровержение) гипотезы, либо *установить надежность оценки,* полученной в результате такого эксперимента или испытания… *Во-вторых,* статистика является основной отраслью прикладной математики, поэтому в выборе принципов и методов мы должны руководствоваться их практическими применениями» [1, с. 7].

В практике спортивных исследований статистические методы позволяют решить следующую типовую задачу: две группы спортсменов, занимающиеся одним и тем же видом спорта, тренировались с использованием различных (традиционной и экспериментальной) методик. Для оценки сравнительной эффективности экспериментальной методики необходимо выявить отсутствие или наличие статистических различий между группами по какому-либо признаку. Задача решается с помощью стандартных статистических методов, но с учётом того, что сферой приложения является область спорта.

Целью данной работы является обоснование выбора статистического метода, адекватного поставленной задаче. Для решения задачи выдвигаются две противоречащие друг другу статистические гипотезы:

– нулевая гипотеза H_0: по рассматриваемому признаку нет существенных различий между группами;

– конкурирующая (альтернативная) гипотеза H_1: есть существенные различия по рассматриваемому признаку между данными группами.

При проверке сформулированных гипотез может возникнуть одна из

1. Предполагается, что сопоставляемые наборы результатов являются выборками из генеральной совокупности, распределенной по некоторому закону, модель которого с характеризующими его параметрами выбрана нами предварительно. Чаще всего в качестве такой модели принимают нормальное распределение, встречающееся наиболее часто в научно-естественных исследованиях. Гипотеза о согласии распределения рассматриваемой совокупности выбранной модели распределения не отвергается, если в соответствии с правилами проверки статистических гипотез будет доказано, что при имеющемся наборе экспериментальных данных гипотетическое распределение может иметь место. Если гипотеза о согласии не отвергнута, мы, используя критерии, основанные на известных параметрах распределения (потому критерии и называются *параметрическими*), принимаем решение о справедливости одной из сформулированных выше гипотез H_0 или H_1.

2. Если же по полученному набору экспериментальных данных получается результат, который при истинности выдвинутой нулевой гипотезы маловероятен, гипотеза о согласии отвергается [2]. Границей маловероятности (невозможности) обычно считают $\alpha = 0,05$ или, что бывает реже, 0,01. Эта граница называется уровнем значимости. В случае отклонения данной гипотезы используют критерии, в расчетные формулы которых не входят никакие параметры, а только частоты или ранги. Поэтому такие критерии называют *непараметрическими*.

Таким образом, в качестве **задач нашей работы** мы выбрали следующие:

1. Выбор критериев, которые было бы лучше использовать в практике статистической обработки результатов спортивных измерений: параметрических или непараметрических. В общем случае, если выполняются условия применения параметрических методов (например, нормальное распределение генеральных совокупностей и равенство их дисперсий), то их и следует применять независимо от специфики области приложения. Во всех остальных случаях рекомендуются непараметрические критерии [3].

2. Проверить на практике применение выбранного критерия для решения задачи оценки результатов научного исследования в биомеханике, чтобы убедиться в его простоте, наглядности и эффективности.

Основной материал исследования. Одним из авторов данной статьи было высказано предположение, что параметрические критерии могут использоваться только в 25 % случаев [4, 5]. Позже на основании большого количества расчетного материала был подсчитан процент случаев, когда мы можем с полным правом использовать параметрические критерии. Оказалось, что такое право мы имеем в 27 % случаев. Получается, что в остальных 73 % случаев нам остается полагаться на непараметрические критерии.

Подробное сопоставление возможностей и ограничений параметрических и непараметрических критериев приведено в работе Е. В. Сидоренко [6]. Автор отмечает, что «параметрические критерии могут оказаться несколько более мощными, чем непараметрические, но только в том случае, если признак измерен по интервальной шкале и нормально распределен» [6, с. 29]. Далее отмечается, что «мощность критерия – это его способность выявлять различия, если они есть. Иными словами, это его способность отклонить нулевую гипотезу об отсутствии различий, если она неверна. Мощность критерия определяется эмпирическим путем. Одни и те же задачи могут быть решены с помощью разных критериев, при этом обнаруживается, что некоторые критерии позволяют выявить различия там, где другие оказываются неспособными это сделать, или выявляют более высокий уровень значимости различий. Возникает вопрос: а зачем же тогда использовать менее мощные критерии? Дело в том, что основанием для выбора критерия может быть не только мощность, но и другие его характеристики, а именно: а) простота; б) более широкий диапазон использования (например, по отношению к данным, определенным по номинативной шкале или по отношению к большим n); в) применимость по отношению к неравным по объему выборкам; г) большая информативность результатов» [6, с. 32].

Большинство непараметрических критериев удовлетворяют перечисленным характеристикам. «По сравнению с параметрическими критериями они ограничены лишь в одном, – с их помощью невозможно оценить взаимодействие двух или более условий или факторов, влияющих на изменение признака. Эту задачу может решить только дисперсионный двухфакторный анализ» [6, с. 29].

Г. Хан и С. Шапиро [7] отмечают, что в статистических исследованиях часто злоупотребляют гипотезой о нормальном законе распределения. Зачастую она выдвигается в результате недостаточной осведомленности исследователя в учении о функциях распределения. При этом, как отмечалось выше, гипотеза не отвергается, если имеющиеся наблюдения не дают результат, являющийся маловероятным при истинности выдвинутой гипотезы. Однако при недостаточном объеме выборочных данных полученное распределение можно столь же успешно соотнести с любым другим законом распределения, например, равномерным. А если в соответствии с принятым нормальным законом распределения провести экстраполяцию результатов за границы области исследования, это может привести к большим ошибкам [8].

Вся приведенная нами информация и сопутствующие рассуждения подводят к мысли о предпочтительности непараметрических критериев. Однако в общем случае решение о выборе критерия исследователь принимает, исходя из условий задачи и имеющихся рекомендаций. Рассмотрим несколько вариантов рассматриваемых задач, которые могут решаться в практике спортивных исследований.

1. Бывают случаи, когда измерение признака в интервальной шкале невозможно, например, когда исследуемый результат выражен в ранговой шкале. Это может быть место, занятое на соревнованиях по гимнастике, фигурному катанию, единоборствам, прыжкам в воду и т.д. В этом случае можно использовать лишь *непараметрические* критерии.

2. Признак измерен в шкале интервалов или отношений. При этом обе выборки или хотя бы одна из них взяты из совокупности, распределенной по закону, отличающемуся от нормального. В таком случае также могут быть использованы только *непараметрические* критерии.

3. Исследуемый признак измерен в шкале интервалов или отношений, обе выборки взяты из нормально распределенных генеральных совокупностей. В этом случае могут быть использованы *параметрические* критерии.

Стоит отметить, что даже в последнем случае использовать параметрические критерии следует с большой осторожностью, поскольку гипотеза о нормальном распределении может быть не отвергнута только лишь по причине недостаточного количества исследуемого материала. При этом у нас нет уверенности, что с увеличением объема выборок не получится результат, вынуждающий нас отвергнуть проверяемую гипотезу. В спорте высоких достижений провести проверку гипотез на выборках большого объема удается далеко не всегда из-за незначительного количества спортсменов высокого класса.

Таким образом, возникает дилемма: использовать более мощный параметрический критерий, в правомерности использования которого мы не вполне уверены, или менее мощный непараметрический критерий, не имеющий ограничений своего применения. Наши рассуждения более склоняют нас к выбору второго варианта, поскольку в случае ошибочности принятия нами гипотезы о нормальном распределении параметрический критерий окажется менее мощным и приводящим с большой вероятностью к ошибочным решениям; но если даже менее мощный непараметрический критерий приведет нас к решению об отклонении нулевой гипотезы об отсутствии различий, более мощный параметрический критерий неизбежно даст такой же результат.

Рассмотрим *пример* использования непараметрического критерия для сравнения двух количественно выраженных признаков.

Две группы студенток непрофильного вуза по 15 человек занимались на протяжении учебного года комплексом гимнастических упражнений, направленных на увеличение амплитуды изменения суставных углов в тазобедренных суставах. Одна группа (контрольная) занималась по традиционной методике, другая группа (экспериментальная) – по новой методике, предположительно более эффективной по сравнению с традиционной. По окончании упомянутого периода времени было решено проверить сравнительную эффективность новой методики, используя тестовое упражнение на гибкость – наклон вперед из положения сидя. Изначально обе группы по упомянутому признаку были однородными. После тренировок по традиционной и новой методикам были получены результаты, представленные в таблице 1.

Таблица 1

**Результаты измерения в тестовом упражнении на гибкость (в см)
в контрольной и экспериментальной группах по окончании периода тренировок**

№ п/п (i)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i (контр.)	17	-15	13	10	19	16	2	8	10	12	-10	-9	0	-3	-13
y_i (эксп.)	19	11	22	18	21	-13	-15	-9	-11	-12	-8	-15	-8	-9	-7

Нетрудно определить, что средний результат тестового упражнения в контрольной группе получился 3,80 см, а в экспериментальной группе -1,07 см. Однако можно ли при такой разнице результатов считать, что в экспериментальной группе получен существенно лучший показатель, чем в контрольной группе? Иначе говоря, нам надо проверить следующие гипотезы:

– нулевую $H_0: \bar{x}^{ген.}_{контр.} = \bar{y}^{ген.}_{эксп.}$, то есть различия между средними значениями признака в группах несущественны;

– конкурирующую $H_1: \bar{x}^{ген.}_{контр.} > \bar{y}^{ген.}_{эксп.}$, то есть в экспериментальной группе результат можно считать значительно меньшим по сравнению с контрольной группой.

Проверку гипотез будем проводить на уровне значимости 0,05 с использованием непараметрического критерия Манна–Уитни. Данный критерий в ряде случаев является очень удобным, поскольку может сравнивать средние значения выборок разного объема, а также малых выборок с объемом от трех единиц. Одна из сравниваемых выборок может иметь даже две единицы, при условии, что другая выборка содержит не менее чем пять единиц.

Порядок применения критерия Манна–Уитни следующий.

1. Составляется единый ранжированный ряд из обеих сопоставляемых выборок, расставив их элементы по степени нарастания признака и приписав меньшему значению меньший ранг. Общее количество рангов получится равным $N = n_1 + n_2$, где n_1 – количество единиц в первой выборке, а n_2 – количество единиц во второй выборке. В нашем примере $n_1 = n_2 = 15$; $N = 30$.

2. Единый ранжированный ряд делится на два ряда, состоящие соответственно из единиц первой и второй выборок. Подсчитывается отдельно сумма рангов, пришедшихся на долю элементов первой выборки, и отдельно – на долю элементов второй выборки. Определяется большая из двух ранговых сумм (T_x), соответствующая выборке с n_x единиц.

3. Определяется значение U-критерия Манна-Уитни по формуле:

$$U = n_1 + n_2 + \frac{n_x \times (n_x + 1)}{2} - T_x .$$

4. По таблице для избранного уровня статистической значимости определяется критическое значение критерия для данных n_1 и n_2. Критерий Манна–Уитни определяет степень совпадения количественных результатов сравниваемых выборок. Чем больше полученное значение критерия, тем более значения совпадают. Если полученное значение U меньше табличного или равно ему, то признается наличие существенного различия между уровнем признака в рассматриваемых выборках (принимается альтернативная гипотеза). Если же полученное значение U больше табличного, принимается нулевая гипотеза о несущественности различий. Достоверность различий тем выше, чем меньше значение U, что означает меньшее совпадение полученных результатов.

Порядок расчета критерия Манна–Уитни для нашего примера приведен в таблице 2.

Таблица 2

Расчет критерия Манна–Уитни

№ п/п	x_i (контр.)	$R_{\text{контр.}}$	y_i (эксп.)	$R_{\text{эксп.}}$
1	-15	2		
2			-15	2
3			-15	2
4	-13	4,5		
5			-13	4,5
6			-12	6
7			-11	7
8	-10	8		
9	-9	10		
10			-9	10
11			-9	10
12			-8	12,5
13			-8	12,5
14			-7	14
15	-3	15		
16	0	16		
17	2	17		
18	8	18		
19	10	19,5		
20	10	19,5		
21			11	21
22	12	22		
23	13	23		
24	16	24		
25	17	25		
26			18	26
27	19	27,5		
28			19	27,5
29			21	29
30			22	30

$\sum R_{контр.} = 251 = T_x$ $\qquad\sum R_{эксп.} = 214$

Наблюдаемое значение критерия Манна–Уитни получаем:

$$U = 15 \times 15 + \frac{15 \times (15 + 1)}{2} - 251 = 94.$$

Полужирным шрифтом обозначены одинаковые значения, которым присваивается ранг равный среднему арифметическому значению всех рангов, занимаемых этими значениями.

По таблице критических значений критерия Манна–Уитни на уровне значимости $\alpha = 0,05$ для объемов выборок $n_1 = n_2 = 15$ получаем $U_{набл.} = 94$.

Так как наблюдаемое значение 94 больше критического 72, мы не отвергаем нулевую гипотезу о том, что различие между средними значениями изучаемого показателя в контрольной и экспериментальной группах является несущественным.

Выводы. Выбирая критерий для статистической обработки результатов измерений в спортивных исследованиях, в том числе и в биомеханике, необходимо учитывать прежде всего корректность использования тех или иных критериев в каждом конкретном случае. Стоит заметить, что непараметрические критерии, в отличие от параметрических, не имеют ограничений в применении и могут быть рекомендованы для статистической обработки, поскольку в области спорта выборки как правило бывают небольшими, а результаты зачастую измеряются в ранговой шкале (занятые места, баллы и т.д.).

Пример применения непараметрического критерия Манна-Уитни для сравнения показателей двух непарных выборок показал простоту данного критерия, его наглядность и эффективность. В приведенном случае совпадение результатов в обеих группах оказалось значительным, что не дало основания для отвержения нулевой гипотезы, то есть существующее различие в средних значениях выполнения тестового задания в контрольной и экспериментальной группах нельзя считать существенным.

Перспективы дальнейших разработок. В дальнейшем мы предполагаем включить в образовательный процесс на кафедре биомеханики Белорусского государственного университета физической культуры обязательное изучение непараметрической статистики с целью ее применения для анализа учебных заданий, а также для научно-исследовательской работы студентов, магистрантов и аспирантов.

Использованные источники

1. Питмен Э. Основы теории статистических выводов: Пер. с англ. / Э. Питмен. – М. : Мир, 1986. – 104 с.
2. Шупляк В. И. Математическая статистика: курс лекций / В. И. Шупляк. – Минск: РИВШ, 2011. – 228 с.
3. Шторм Р. Теория вероятностей. Математическая статистика. Статистический контроль качества / Р. Шторм. – М. : Мир, 1970. – 368 с.
4. Волков Ю. О. Непараметрические критерии статистического оценивания в спортивных исследованиях / Ю. О. Волков, В. К. Пономаренко, С. Л. Рукавицына // Materialy IV Miedzynarodowey konferencji «Wykształcenie I nauka bez graniz – 2008, 07 – 15 grudnia 2008 roku. – Przemysl : Nauka I studia, 2008. – S. 58 – 62.
5. Пономаренко В. К. Применение непараметрических критериев для проверки статистических гипотез в спортивных исследованиях / В. К. Пономаренко, Ю. О. Волков // Научное обоснование физического воспитания, спортивной тренировки и подготовки кадров по физической культуре, спорту и туризму : материалы XV Междунар. науч. сессии по итогам НИР за 2016 год, посвященной 80-летию университета, Минск, 30 марта – 17 мая 2017 г. : в 4 ч. / Белорус. гос. ун-т физ. культуры ; редкол. : Т. Д. Полякова (гл. ред.) [и др.]. – Минск : БГУФК, 2017. – Ч. 2. – С. 107 – 111.
6. Сидоренко Е. В. Методы математической обработки в психологии / Е. В. Сидоренко. – СПб : ООО «Речь», 2000. – 350 с.
7. Хан Г. Статистические модели в инженерных задачах / Г Хан, С. Шапиро. – М. : Мир, 1969. – 395 с.
8. Налимов В. В. Теория эксперимента / В. В. Налимов. – М. : Наука. Главная редакция физико-математической литературы, 1971. – 208 с.

Volkov Yu., Ekimov V.., Shinder M.

THE CRITERIA FOR ESTIMATING THE RESULTS OF EXPERIMENTAL RESEARCH IN BIOMECHANICS

Professor E. Pitman, a prominent Australian mathematician and specialist in mathematical statistics, came to the conclusion that the theory of statistical inference requires two points to be taken into account. First, to formulate principles that will help statisticians to assess the quality of experimental data or tests in support (or refutation) of the hypothesis, or to determine the reliability of the evaluation obtained as a result of such an experiment or test. Secondly, statistics is the main branch of applied mathematics, therefore, when choosing the principles and methods we must be guided by their practical applications.

In the practice of sports research, statistical methods allow to solve the following routine problem: two groups of athletes engaged in the same sport were trained using various (traditional and experimental) techniques. To assess the comparative effectiveness of the experimental methodology, it is necessary to identify the absence or presence of statistical differences between groups on some basis. The problem is solved using standard statistical methods, but taking into account the fact that the area of application lies in the field of sports.

To solve this problem, two contradictory statistical hypotheses are put forward:

– null hypothesis H_0: there are no significant differences in the feature under consideration between groups;

– competing (alternative) hypothesis H_1: there are significant differences in the feature under consideration between these groups.

When testing the formulated hypotheses, one of two possible situations may arise.

It is assumed that the compared sets of results are samples from the entire assembly, distributed according to a certain law, the model of which being previously selected and characterized by certain parameters. Typically, such a model is assumed to be a normal distribution, which is most frequently encountered in the scientific and natural research. The hypothesis of the distribution concordance of the considered assembly of the selected distribution model is not rejected if, in accordance with the rules for checking statistical hypotheses, it will be proved that, the available set of experimental data allows a hypothetical distribution. If the hypothesis of concordance is not rejected, we make a decision whether one of the hypotheses H_0 or H_1 formulated above is valid, using the criteria based on known distribution parameters (therefore the criteria are called parametric).

If the result obtained on the basis of the set of experimental data gathered is unlikely, the proposed null hypothesis being true, the hypothesis of concordance is rejected. The boundary of improbability (impossibility) is usually considered to be $\alpha = 0.05$ or, which happens more rarely, 0.01. This boundary is called the significance level. In the case of rejection of this hypothesis, the criteria used are based on calculation formulas with no other parameters but frequencies or ranks. Therefore, such criteria are called nonparametric.

Key words: *statistics, hypotheses, criteria, experimental studies, biomechanics.*

Стаття надійшла до редакції 20.09.2018 р.