DOI 10.26773/smj.210929

SportMont

ORIGINAL SCIENTIFIC PAPER

Formation of the Knowledge and Skills to Apply Non-Parametric Methods of Data Analysis in Future Specialists of Physical Education and Sports

Nataliia Byshevets[1], Olena Iakovenko[1], Olha Stepanenko[1], Kostyantyn Serhiyenko[1], Yuriy Yukhno[1], Nataliia Goncharova[1], Nataliia Blazhko[1], Mariia Kolchyn[1], Hanna Andriyenko[1], Nadiya Chyzhevska[1] and Taras Blystiv[2]

[1]National University of Ukraine of Physical Education and Sport, Innovation and Information Technologies in Physical Culture and Sports Department, Kyiv, Ukraine, [2]National University of Ukraine of Physical Education and Sport, Kyiv, Department of Health, Fitness and Recreation, Kyiv, Ukraine

Abstract

Application of information technologies in the process of research in the field of physical education and sport, sports medicine allows to effectively use mathematical and statistical methods in scientific research and teaching. We have highlighted the main stages of practical training aimed at the formation of practical skills in the use of non-parametric methods for estimating and analyzing statistical hypotheses in sports-pedagogical and biomedical research among the students in the field of physical education and sports. It was noted that the role of qualitative statistical processing of research results is growing. It is important to form the skills the use of the methods of mathematical statistics of future specialists in various fields. The article presents a systematic process for the formation practical skills and abilities to use non-parametric procedures for statistical analysis of sports pedagogical data on the example of the Mann-Whitney U-test. There are the following stages of the training: announcement of the theme, goals, tasks; actualization of core knowledge; formation of theoretical skills. Article contains stages the presentation of the algorithm for the comparative analysis of independent sample data that does not follow the normal distribution law. It is shown calculation of a working example of the Mann-Whitney U-test using MS Excel in group using a interactive whiteboard with a detailed explanation of the steps; individual solving of the practical tasks; knowledge check.

Keywords: *formation, research, statistics, analysis, criterion, law, distribution*

Introduction

The scientific basis for obtaining reliable experimental results is the mathematical-statistical processing of empirical data, and the application of information technologies in the process of research in the field of medicine and sports opens up significant prospects for the application of mathematical-statistical methods to a wide range of researchers (Byshevets, Synihovets, & Oliynyk, 2011; Byshevets, 2017).

Analyzing the results of sport-pedagogical and biomedical experiments, the researcher often faces the task of testing statistical hypotheses (Shynkaruk, 2012; Stroganov et al., 2020; Tukaiev et al., 2020; Tretiak et al., 2020). The Student t-test is especially popular among scientists. However, as it is well known, the use of parametric criteria needs the fulfillment of a number of conditions. Researchers should pay attention to the condition of normal distribution of each of the compared samples.

Unfortunately, as Leonov (2007) points out, the Student's t-test is often used without justification, which negates the perfect experimental work and gives rise to doubts about the correctness of the declared conclusions.

The ordinal variables data does not follow the normal distribution law while analyzing sports-pedagogical and biomedi-

Correspondence:

N. Byshevets
National University of Ukraine of Physical Education and Sport, Kyiv, Innovation and Information Technologies in Physical Culture and Sports Department, Kyiv 03150, 1, Fizcultury str., Kyiv, Ukraine
E-mail: bishevets@ukr.net

cal data, especially when it comes to small sample sizes. In such cases, scientists use non-parametric methods for estimating and analyzing statistical hypotheses. They are computational procedures that do not use the distribution parameters, but use the results of ranking or calculating the values of the test feature (Lupan, 2010; Lang, 2016). At the same time, while retaining most of the information on the distribution, the use of non-parametric statistical methods eliminates the need to know what the distribution is, and there is no need for equality of variances (Glantz, 1998).

The Mann-Whitney U-test is used in order to identify the difference between the levels of the studied features of two independent samples. It helps to reveal the differences between the samples by estimating the width of the total area for both samples (Mann, 1948).

Purpose is to identify the stages of practical training aimed at developing practical skills in the use of non-parametric methods for estimating and analyzing statistical hypotheses in sports-pedagogical and biomedical research among students in the field of physical education and sports.

Methods

Used methods include the analysis of scientific and methodical literature, Internet data and mathematical statistics methods, including data processing.

Results

In the course of the study, we distinguished the stages of formation readiness to use non-parametric criteria in sports-pedagogical and biomedical research on the example of using the Mann Whitney U-test for students of physical education and sports training.

We give an example of practical training aimed at the formation of theory and practical abilities realize a comparative analysis of independent sample data that are not followed the normal distribution law, using MS Excel.

After informing the topic and goals, we define the tasks must be solved during the practical training.

1. To establish the purpose and accuracy of the Mann-Whitney U-test.

2. To learn to compare sample data using the Mann-Whitney U-test.

At the first stage, there is an actualization of students' basic knowledge. During this stage, they are surveyed, where the question is chosen in such a way as to reveal the final knowledge on the themes "Normal law of distribution and its application", "Testing statistical hypotheses", "Statistical criteria". This stage is intended to prepare students for a new theme in such a way that the formation of new theoretical knowledge takes place based on previously learned and deepened them.

At the second stage, we provide students with basic information and concepts. The purpose of this stage is the formation of theoretical knowledge of students about the features of the U-criterion and the conditions for its use.

The Mann-Whitney U-test is a non-parametric alternative to the Student's t-test for independent samples, which is commonly used to process the data obtained in the field of physical education and sports. The advantage of this criterion is that its use does not require assumptions about the normal distribution and the same variances.

The method is based on determining whether a sufficiently small area of values between two variational series that intersect. The smaller the value of the criterion, the more likely it is that the differences between the parameter values in the samples are significant.

At this stage, we stress the features of the U-criterion and the conditions for its use.

The Mann-Whitney U-test is a non-parametric test, therefore, unlike Student's t-test; it does not require a normal distribution of compared populations.

It is necessary that the data be measured at least in an ordinal scale.

The U-criterion is suitable for comparing small samples: in each of the samples, there must be at least 3 trait features.

It is allowed that one of the samples had only two features, but another one must have at least five features.

The condition for applying the Mann-Whitney U-test is the absence of matching features in the compared groups (all numbers are different) or the minimum number of such matches.

Also at this stage, we form a hypothesis in the following way: null hypothesis H0 {Character level in sample 2 is not lower than the characteristic level in sample 1}, and alternative to it is H1 {Character level in sample 2 is lower than the characteristic level in sample 1}

At the third stage, we inform the algorithm for the implementation of the comparative analysis of independent sample data that are not followed the normal distribution law, using MS Excel with a detailed explanation and group discussion of each step.

1. We make the initial data in the MS Excel table and make it in a convenient form.

2. We make a logical assumption about the presence of cause-and-effect relations between the studied features X and Y and make hypotheses: H_0 {the distribution of the sign of choice 1 corresponds to the distribution of the sign of choice 2}; H_1 {the distribution of the sign of choice 1 does not correspond to the distribution of the sign of choice 2}.

3. Perform data ranking, treat as one array.

4. We calculate the U-criterion according to the formula:

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

where n1, n2 – number of elements in samples, nx – the number of elements of a larger rank sums; – larger of two rank sums.

5. As criterion statistics, we use the value of U, the critical region of which is determined using the special Mann-Whitney distribution. That is, the critical value of the U-criterion is found using the statistical table, where the significance level α is taken 0.05, and for ni - the sample sizes i=1, 2.

6. We compare the calculated and critical value of U. We conclude: if U is critical exceeds U empirical, then the statistically significant differences between the studied groups based on feature, that is being studied, are not confirmed (Lupan, 2014).

Important! The number at the intersection of the sample size and the smallest sample is the critical value of the Mann-Whitney U-test.

The next fourth stage, aimed at developing students' practical skills and abilities to apply non-parametric the Mann-Whitney U-test for processing the results of sports-pedagogical and biomedical research, involves the instructions of an example solving a task. At this stage, there is a working example of solving the task and interpreting the obtained results. During this stage, the clarification of individual concepts and a group discussion of the result.

The stage of formation of practical skills is very important, so it is necessary to ask students to consider a practically oriented task, and also it is reasonable to present a step-by-step realization of the method of analysis. Students should also be encouraged to participate in a collective discussion on the relevance of using nonparametric methods of data analysis in physical culture and sports. Also in the demonstration process the steps of problem solving are clarified, the difficulties encountered by the students are identified.

So, we offer the students a professionally oriented task for consideration.

Example. There are indicators of task performance of the two study groups. Prove that the task performance of second-year students exceeds the task performance of first-year students.

The demonstration of the steps for solving the task is carried out using an interactive whiteboards.

Step 1. Run MS Excel, form a table and enter the output data into it.

Step 2. We form hypotheses: H_0 {task performance of first-year students is notlower than task performance of second-year students}; H_1 {task performance of first-year students is lower than task performance of second-year students}.

Give the significance level $\alpha = 0.05$ ($p<0.05$).

Step 3. We compile a common ranked series using the standard function RANK.AVG, where a smaller value is assigned a lower rank. In addition, in case of coincidence of values, the arithmetic mean of consecutive values of ranks is assigned to each of them.

Create cells n1 and n2, in which using the standard function COUNT we calculate the number of elements in each sample.

Create cells Rx sum and Ry sum, in which, using the standard SUM function, we calculate the sum of the ranks of each sample.

Calculate T_x. To do this, using the function IF, compare the values of ΣR_x and ΣR_y and choose larger of them.

Calculate nx. To do this, using the function IF, compare the values of ΣR_x and ΣR_y and choose number of elements larger of a rank sums. In the process of this step, we focus on the fact that when introducing formulas it is convenient to rely on the following considerations (Figure 1).

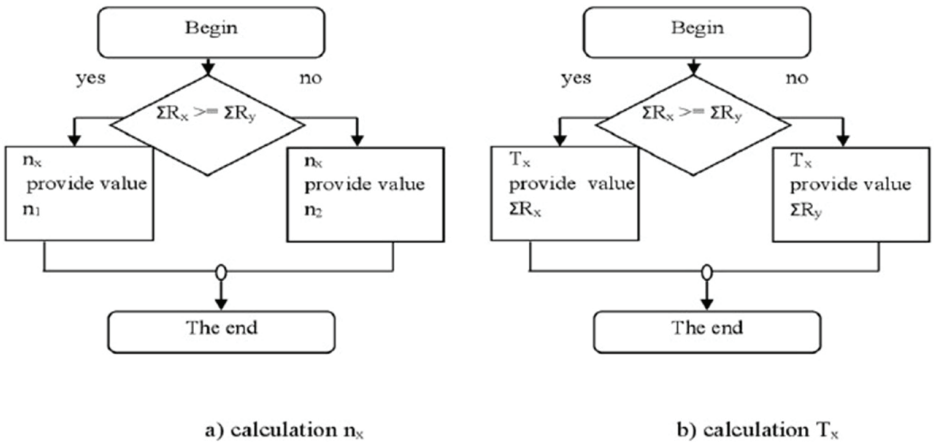

FIGURE 1. Algorithm for calculating the number of elements of a larger sample of n_x and the larger of the two rank sums T_x.

Step 4. Calculate the U-test by the specified formula.

Step 5. According to the table of critical values, we calculate U critical with sample sizes n1=20 i n2=16 at the significance level α=0.05. We got.

Step 6. We make the following conclusions: comparing the critical and empirical value of U, we are convinced that $U_{emp}>U_{kr}$. Consequently, it is no statistically significant differences between the task performance of the students of the first

C	D	E	F	G	H
Age		R_x	R_y		
I	II				
838	701	27	11		
643	645	5	6		
583	702	2	12		
709	803	14	24		
538	800	1	22,5		
714	710	16	15		
778	680	20	7	p	0,05
903	820	33	25		
741	681	18	8	n_1	20
1020	700	34,5	10	n_2	16
1040	730	36	17		
1020	620	34,5	3	ΣR_x	419
799	800	21	22,5	ΣR_y	247
685	901	9	32		
626	703	4	13	n_x	20
830	742	26	19	T_x	419
848		28			
854		30		U	111
882		31		$U_{cr}(0,05;20;16)$	98
852		29		result	H0

=RANK.AVG(D3;$C$3:$D$22;1)
=COUNT(C3:C22)
=SUM(E3:E22)
=IF(H11>H12;H11;H12)
=IF(H14>H15;H14;H15)
=H11*H12+H17*(H17+1)/2-H18
=IF(H20>H21;"H0";"H1")

FIGURE 1. The implementation of the calculation of the criterion of Mann-Whitney in MS Excel

year students and second year students (p=0.164> 0.05). The results of the task are shown in the figure (Figure 2).

During the implementation of the fifth stage of practical training, students are offered individual tasks, including test questions and practice-oriented tasks. Students aim this stage at identifying the level of mastery of practical work.

The purpose of the last stage is the operational control of the students' achievement. At this stage, an assessment of the results of passing tests on test questions is carried out; the result of solving the task is checked.

Discussion

Currently, the increasing role of statistical data processing requires the formation of students' skills and abilities to apply the methods of mathematical statistics. Now, to improve the system of training future specialists in physical education and sport, issues related to the formation of theoretical knowledge and practical skills to apply certain methods of data analysis are covered. In particular, an algorithm for analyzing personal data in sports pedagogical research was proposed (Denysova et al., 2012; Byshevets et al., 2019; Denysova et al., 2020), ways to automate the calculation of analysis results were proposed (Synihovets, 2011) and a methodology of preparing students of higher educational institutions for physical education to use the method of expert assessments was presented (Byshevets, 2018; Byshevets et al., 2019).

However, according to scientists (Leonov, 2007), the statistical processing of experimental data remains the weakest point of research in the field of medicine, biology, ecology, due to insufficient equipment of statistical knowledge and insignificant experience enrolled in these areas. Own experience suggests that students of the field of physical education and sports, unfortunately, face significant difficulties in studying the disciplines related to the statistical training of future specialists. Indeed, in many cases, the level of mathematical training of students often does not provide opportunities to master educational material, designed for a solid knowledge base (Byshevets, 2017; Kyslenko et al., 2020). Therefore, there is an urgent need to show the most accessible methods and techniques for automating the process of solving professionally oriented tasks using computer systems and programs for data analysis. It should be noted that the use of application programs and data analysis packages makes available statistical processing of empirical data of any complexity. Moreover, it is known that a positive experience encourages students to further research and practical developments in this direction.

At the same time, we agree with Toropova (2017), teachers should pay more attention not only to the formation of statistical knowledge and skills at students of non-core higher educational institutions, but also to focus on training students to use this knowledge to statistically process the research results doing course project, diploma project. This approach allows students in practice to realize the role and place of statistical data processing of experimental studies and to acquire the skills obtained through systematic recapitulation.

Acknowledgements
There are no acknowledgements.

Conflict of Interest
The authors declare that there are no conflicts of interest.

Received: 27 November 2020 | **Accepted:** 22 January 2021 | **Published:** 01 September 2021

At the same time, scientists note the need to apply distance-learning technologies when studying a course aimed at developing skills for statistical data processing of students who are at individual form of teaching (Krivtsova, 2016; Denysova et al., 2018; Imas et al., 2018). From our point of view, such an approach is appropriate for students of higher educational institutions in the field of physical education, since a significant part of the students are active athletes who are not able to attend classes regularly. Hence, the development of a distance course aimed at developing practical skills and abilities to carry out statistical data processing is a priority for improving the system of training future specialists in physical education and sports.

Conclusions

The role of statistical analysis of research results in various fields of knowledge is increasing, which has led to great interest at scientists to improve the statistical training of future specialists.

Statistical data processing causes difficulties both for students of non-core higher educational institutions and for researchers in physical culture and sports, medicine, biology, and ecology. This situation leads to errors that occur during the analysis of empirical data.

The most promising direction to optimize the educational process, aimed at the formation of theoretical knowledge, skills and abilities of statistical processing of empirical data is the use of modern information technologies at all stages of training. This opens up the possibility for students to carry out a qualitative analysis of research results without taking into account prior mathematical grounding. At the same time, the positive experience of solving professionally oriented tasks contributes to increasing the motivation of future specialists in physical education and sports to use learned methods in research practice.

A lot of active athletes are students in the field of physical culture and sport, as a result, they have to study at the individual form of teaching. Therefore, it is important to apply distance-learning technologies, including in the formation of student's statistical knowledge and skills.

The Mann-Whitney U-test is a non-parametric analogue of Student's t-test. However, unlike the latter, its use does not require that the compared samples follow the normal distribution law and there is no need for equality of the variances.

As a result of the study, a systematic process has been developed for students to form practical skills in using non-parametric procedures for statistical analysis of sports pedagogical data using the Mann-Whitney U-test. There are the following stages: announcement of the theme, goals, tasks; actualization of core knowledge; formation of theoretical skills; the presentation of the algorithm for the comparative analysis of independent sample data that does not follow the normal distribution law; implementation of the calculation of the Mann-Whitney U-test using MS Excel; independent solving of practical tasks; knowledge check.

References

Byshevets, N. G., Serhiyenko, K. M., & Golovanova, N. L. (2018). Preparing students of higher educational institutions in the field of physical education to use the method of expert assessments. *Theory and methods of physical education, 1*, 18–35.

Byshevets, N., Denysova, L., Shynkaruk, O., Serhiyenko, K., Usychenko, V., Stepanenko, O., & Syvash, I. (2019). Using the methods of mathematical statistics in sports and educational research. *Journal of Physical*

Education and Sport, *19*(3), 1030–34. doi:10.7752/jpes.2019.s3148

Byshevets, N., Shynkaruk, O., Stepanenko, O., Gerasymenko, S., Tkachenko, S., Synihovets, I., Filipov, V., Serhiyenko, K., & Iakovenko, O. (2019) Development skills implementation of analysis of variance at sport-pedagogical and biomedical researches. *Journal of Physical Education and Sport*, *19*(6), 311, 2086–90. doi:10.7752/jpes.2019.s6311

Byshevets, N., Synihovets, I., & Oliynyk, R. (2011). Comparing analysis in scientific and sports activities. *Bulletin of the T.H.Shevchenko National University "Chernihiv Colehium"*, *86*(1), 23–8.

Byshevets, N.G. (2017). Experience of application of modern means of education in practical lessons on the probability theory. *Information technologies in education*, 2 (31), 95–108.

Denysova, L. Shynkaruk, O. & Usychenko, V. (2018) Cloud technologies in distance learning of specialists in physical culture and sports. *Journal of Physical Education and Sport*, *18*(1), 66, 469-72. doi:10.7752/jpes.2018.s166

Denysova, L., Byshevets, N., Shynkaruk, O.,Imas, Ye., Suschenko, L., Bazylchuk, O., Oleshko, T., Syvash, I., & Tretiak, O. (2020). Theoretical aspects of design and development of information and educational environment in the system of training of masters in physical culture and sport. *Journal of Physical Education and Sport, 20*(1), 324– 30. doi:10.7752/jpes.2020.s1045

Denysova, L., Usychenko, V., & Byshevets, N. (2012). The algorithm of questionnaire data analysis in sports and pedagogical researches. *Pedagogics, psychology, medical-biological problems of physical education and sport, 1*, 56–60.

Glantz, S. (1998). *Medical and biological statistics.* Praktika (Practice).

Imas, E., Shynkaruk, O., Denysova, L., Usychenko, V., & Kostykevich, V. (2018) Physical and mental human health in the contemporary information environment. *Journal of Physical Education and Sport*, *18*(3), 261, 1791-95. doi:10.7752/jpes.2018.03261

Krivtsova, O. P., & Dmytriienko, O. O. (2018). The possibility of teaching the course «Statistical processing of the data» by means of Moodle. *Physical &mathematical education*, *2*(16), 63-6.

Kyslenko, D., Shynkaruk, O., Kryshevych, O., Yukhno, Yu., Kopotun, I., Bondarenko, V., & Holovanova, N. (2020). Enhancing physical fitness of future national security specialists of Ukraine using team sports. *Journal of Physical Education and Sport, 20*(1), 54. 378–84. doi:10.7752/jpes.2020.s1054

Lang, T. A., & Sesyk, M. (2016). *How to describe statistics in medicine: a guide for authors, editors and reviewers; transl. from English*. Practical medicine.

Leonov, V. (2007). Errors of statistical analysis of biomedical data. *International Journal of Medical Practice*, *2*, 19-35.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50-60.

Shynkaruk, O. (2012). The concept of formation of a system of training, selection of athletes and their orientation in the process of multi-year perfection. *Pedagogics, psychology and medical and biological problems of physical education and sports, 12*, 144-8.

Stroganov, S., Serhiyenko, K., Shynkaruk, O., Byshevets, N., Denysova, L., Yukhno, Yu., Stepanenko, O., & Ulan, A. (2020). Features of preventive activity at the initial stage of training of many years standing of young basketball players. *Journal of Physical Education and Sport, 20*(1), 452–5. doi:10.7752/jpes.2020.s1066

Synihovets, I. V. (2011). Automatization of expert estimate analysis in physical education and sport. *Bulletin of the T.H. Shevchenko National University "Chernihiv Colehium"*, *91*(2), 98-101.

Toporova, S. I. (2018). Methods of mathematical statistics as a means of professional competence formation of students-ecologists. *The Education and Science Journal, 3*(20), 53-82.

Tretiak, O., Danylo, S., Konovalska, L., Bazylchuk, O., Sushchenko, L., Bazylchuk, V., Shynkaruk, O., & Denysova, L. (2020). Factor structure of interrelation between indices of physical condition and definition of the level of physical readiness of future specialists in higher pedagogical education to act in extreme situations. *Journal of Physical Education and Sport*, *20*(1), 461–8. doi:10.7752/jpes.2020.s1068

Tukaiev, S., Dolgova, O., Van Den Tol, A. J. M., Ruzhenkova, A., Lysenko, O., Fedorchuk, S., Ivaskevych, D., Shynkaruk, O., Denysova, L., Usychenko, V., Iakovenko, O., Byshevets, N., Serhiyenko, K., & Voronova, V. (2020). Individual psychological determinants of stress resistance in rock climbers. *Journal of Physical Education and Sport*, *20*(1), 69, 469–76. doi:10.7752/jpes.2020.s1069