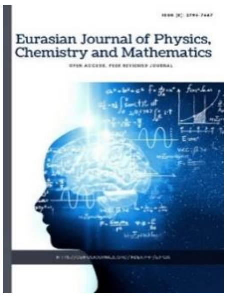

Recommendations for The Selection and Processing of Statistical Methods in Pedagogical Research

Mansurova Nodira Husanbayevna — Masters of the Tashkent State Pedagogical University named after Nizami

Yodgarov Samad Zhurayevich — Associate Professor of the "General Mathematics" Department of Tashkent State Pedagogical University named after Nizami

Rakhmatova Sevinch Kholbekovna — Masters of the Tashkent State Pedagogical University named after Nizami

ABSTRACT

The article considers an algorithm for selecting a statistical criterion for testing statistical hypotheses depending on the type of available data and their distribution. The presented material provides general recommendations on choosing the most appropriate method of data analysis for the most common situations.

Keywords: data type, statistical criteria, distribution

Consider the case when a scale of relations is used for measurements. Suppose there is an experimental group consisting of N people and a control group consisting of M people. Let's assume that as a result of measuring the same indicator using the same measurement procedure, the following data were obtained: : $x = (x_1, x_2, \dots, x_N)$ is a sample for the experimental group and $y = (y_1, y_2, \dots, y_M)$ is a sample for the control group, where x_i is the sample element – the value of the studied indicator for the i–th member of the experimental group, i = 1, 2, ..., N and y_j is the value of the studied indicator for the j–th member of the control group, j = 1, 2, ..., M. Since the measurements were made in the scale of relations, {xi} and {yj} are positive, including, possibly, integers, numbers for which all arithmetic operations make sense. As an example, we will consider the results of measurements of the level of knowledge in the control and experimental groups before and after the experiment (see Table 1) – the number of correctly solved tasks.

For data measured in the scale of relations, it is advisable to use either the Kramer-Welch criterion [6, 7] or the Wilcoxon-Mann-Whitney criterion [1, 4, 6] to test the hypothesis that the characteristics of the two groups coincide. The Kramer-Welch criterion is designed to test the hypothesis that the averages of two samples are equal, the Wilcoxon-Mann-Whitney criterion is more "subtle" (but also more time–consuming) - it allows you to test the hypothesis that two samples are "the same" (including that their averages, variances and all other indicators coincide).

The Kramer-Welch criterion. The empirical value of this criterion is calculated based on information about the volumes of N and M samples x and y, sample averages x and y, and sample variances D_x and D_y of the compared samples according to the following formula:

$$T_{emp} = \frac{\sqrt{MN}\,|\bar{x} - \bar{y}|}{\sqrt{MD_x + ND_y}} \qquad (1)$$

The algorithm for determining the reliability of coincidences and differences in the characteristics of the compared samples for experimental data measured in the ratio scale using the Kramer-Welch criterion is as follows:

1. Calculate the empirical value of the Kramer–Welch criterion for the compared samples using the formula (1).

2. Compare this value with the critical value

$W_{krit} = W_{0,05} = 1{,}96$:

if $W_{emp} \le 1{,}96$, then conclude: "the characteristics of the compared samples coincide at a significance level of 0.05";

if $W_{emp} > 1{,}96$, then conclude "the reliability of the differences in the characteristics of the compared samples is 95%".

Note that we do not consider the question of "which way" the experimental group differs from the control group, that is, whether the studied characteristics have improved or worsened.

The Wilcoxon-Mann-Whitney criterion. This criterion does not operate with the absolute values of the elements of two samples, but with the results of their paired comparisons. Let's take two samples:

$x_1, x_2, \dots, x_N$; $y_1, y_2, \dots, y_M$

and for each element of the first sample $x_1, x_2, \dots, x_N$ we determine the number a_i of elements of the second sample that exceed it in value (that is, the number y_j such that $y_j > x_i$).

The amount

$$\sum_{i=1}^{N} a_i$$

these numbers for all N members of the first sample is called the empirical value of the Mann-Whitney criterion and is denoted by

$$\sum_{i=1}^{N} a_i = U.$$

Let 's define the empirical value of the Wilcoxon criterion:

$$\boldsymbol{W_{emp}} = \frac{\left|\frac{\boldsymbol{NM}}{\boldsymbol{2}} - \boldsymbol{U}\right|}{\sqrt{\frac{\boldsymbol{NM(NM+1)}}{\boldsymbol{12}}}} \qquad (\boldsymbol{2})$$

The algorithm for determining the validity of coincidences and differences for experimental data measured in the scale of relations using the Wilcoxon-Mann-Whitney criterion is as follows:

1. Calculate the empirical value of the Wilcoxon criterion for the compared samples using the formula (2).

2. Compare this value with the critical value

$W_{krit} = W_{0,05} = 1{,}96$:

if $W_{emp} \le 1{,}96$, then we conclude: "the characteristics of the compared samples coincide with the significance level of 0.05";

if $W_{emp} > 1{,}96$, then we conclude "the reliability of the differences in the characteristics of the compared samples is 95%".

The restriction on using the Wilcoxon-Mann-Whitney criterion is as follows: each sample must contain at least three elements, if there are only two elements in one of the samples, then there must be at least five in the second one. It does not matter which sample is considered the first and which is the second, although it is more convenient to calculate the first sample with fewer members.

Methodology determination of the validity of coincidences and differences for experimental data measured in an ordinal scale

Consider the case when an ordinal scale with L different scores is used. The characteristic of the group will be the number of its members who scored one or another point. For the experimental group, the score vector is $n_k = (n_1, n_2, \dots, n_L)$ where n_k is the number of members of the experimental group who received the kth score, k = 1, 2, ..., L. For the control group, the score vector is $m_k = (m_1, m_2, \dots, m_L)$, where m_k, is the number of members of the control group who received the k–th score, k = 1, 2, ..., L. For the numerical example we are considering (L = 3 - "low", "medium" or "high" level of knowledge), the data are given in Table 1.

For data measured on an ordinal scale, it is advisable to use the uniformity criterion χ^2 ("chi" is a letter of the Greek alphabet, the name of the criterion reads: "chi–square") [5], the empirical value of χ^2_{emp} of which is calculated by the following formula:

$$\chi^2_{emp} = M \cdot N \cdot \sum_{i=1}^{L} \frac{(\frac{n_i}{N} - \frac{m_i}{M})^2}{n_i + m_i} \quad (3) .$$

Critical values $\chi^2_{0,05}$ criteriaχ^2 for the significance level of 0.05 are given in Table 1 (statistical tables of critical values of statistical criteria for various levels of significance and various – including large 1 – gradations of the scale of relations can be found, practically, in any textbook on statistical methods, or in special statistical tables [3]).

Table 1. Critical values of the criterion χ^2 for the significance level α= 0.05.

L-1	1	2	3	4	56	6	7	8	9
$\chi^2_{0,}$	3,84	5,99	7,82	9,49	11,07	12,59	14,07	15,52	16,92

The algorithm for determining the validity of coincidences and differences for experimental data measured on an ordinal scale is as follows:

1. Calculate for the compared samples χ^2_{emp} – the empirical value of the criterion χ^2 according to the formula (3).

2. Compare this value with the critical value $\chi^2_{0,05}$:

if $\chi^2_{emp} \leq \chi^2_{0,05}$, then conclude: "the characteristics of the compared samples coincide with with a significance level of 0.05";

if $\chi^2_{emp} > \chi^2_{0,05}$, then conclude "the reliability of the differences in the characteristics of the compared samples is 95%.

The chi-square criterion is applicable provided that for any value of the score in any of the compared samples, at least five of its members received this score, that is: $n_i \geq 5 , m_i \geq 5$, $i = 1,2,3, \dots , L$. In addition, it is desirable that the number of gradations L be at least three. If L = 2, that is, a dichotomous scale is used ("yes" – "no", "decided" – "not decided", etc.), then the Fisher criterion can be applied.

Table 2. The results of measurements of the level of knowledge in the control and experimental groups before and after the experiment.

Level of knowledge	Control group before the experiment (pers.)	Experimental group before the experiment (pers.)	Control group after the experiment (pers.)	Experimental group after the experiment (pers.)
Low	11	8	13	3
Average	16	13	15	12
Tall	9	7	8	13
The amount	36	28	36	28

Let's apply the algorithm for the data from Table 2. First, we calculate the empirical values for the criterion χ^2 using formula (3). For example, here is a calculation.

Parameters of the experimental group (N = 28) after the end of the experiment: : n_1 = 3 n_2 = 12, n_3 = 13 (that is, 3 students demonstrated a "low" level of knowledge, 12 – "average" and 13 – "high"), control group (M = 36): m_1 = 13, m_2 = 15, m_3 = 8. Substituting into the formula (3), we get:

$$\chi^2_{emp} = \frac{1}{36 \cdot 28} \cdot \left(\frac{(11 \cdot 28 - 8 \cdot 36)^2}{11+8} + \frac{(16 \cdot 28 - 13 \cdot 36)^2}{16+13} + \frac{(9 \cdot 28 - 7 \cdot 36)^2}{9+7} \right) = 0{,}03.$$

$$\chi^2_{emp} = \frac{1}{36 \cdot 28} \cdot \left(\frac{(13 \cdot 28 - 3 \cdot 36)^2}{13+3} + \frac{(15 \cdot 28 - 12 \cdot 36)^2}{15+12} + \frac{(8 \cdot 28 - 13 \cdot 36)^2}{8+13} \right) = 6{,}88.$$

Before the experiment : χ^2_{emp}= 0,03, χ^2_{krit}(0,05; 2) = 5,99 , 0,03 < 5,99 $\Longrightarrow$ H_0

After the experiment: χ^2_{emp}= 6,88, χ^2_{krit}(0,05; 2) = 5,99, 6,88 > 5,99 $\Longrightarrow$ H_1.

So, the initial (before the start of the experiment) states of the experimental and

ISSN: 2795-7667

control groups coincide, and the final (after the end of the experiment) – differ. Therefore, it can be concluded that the effect of changes is due to the use of experimental teaching methods.

References

1. 1.Ayvazyan S.A., Mkhitaryan V.S. Applied statistics and fundamentals of econometrics. Moscow: UNITY, 1998. – 1022 p.
2. Ayvazyan S.A., Mkhitaryan V.S. Applied statistics in tasks and exercises. Moscow: UNITY, 2001– - 270 p.
3. Bolshev L.N., Smirnov N.V. Tables of mathematical statistics. Moscow: Nauka, 1983. – 4
4. Tyurin Yu.N., Makarov A.A. Statistical analysis of data on a computer. Moscow: INFRA-M, 1998. – 528 p.
5. Smirnov N.V., Dunin-Barkovsky I.V. Course of probability theory and mathematical statistics for technical applications. M.: Nauka, 1969.
6. Orlov A.I. Econometrics. M.: Exam, 2003. – 576 p.
7. Kramer G. Mathematical methods of statistics.Moscow: Mir, 1975.-648 p.
8. Ismoilov E.O. Pedagogical experiments conducted in the framework of research aimed at developing students' professional competencies on the basis of an integrative approach, their results and statistical analysis // Norwegian Journal of development of the International Science (ISSN 3453-9875) (Journal impact factor 5.02). – Norway, 2022. №83, April, 2022. – p. 22-33. https://www.nor-ijournal.com/wp-content/uploads/2022/05/NJD_83.pdf
9. Ismoilov E.O. Pedagogical experiments on the development of professional competencies of students based on an integrative approach and their results // Periodica Journal of Modern Philosophy, Social Sciences and Humanities (ISSN 2720-4030) (Journal impact factor 5.911). – Poland, 2022. Volume 5, April, 2022. – p. 136-143. https://periodica.org/index.php/journal/article/view/91
10. Ismoilov E.O. Statistical analysis of the results of pedagogical experiments on the development of professional competencies of students on the basis of an integrative approach // International Conference on Developments in Education, Sciences and Humanities. – Livorno, Italy. May 2nd - 3rd 2022. – p. 193-203. https://econferencezone.org/index.php/ecz/article/view/714