

# METHODS FOR NONPARAMETRIC STATISTICS IN SCIENTIFIC RESEARCH. OVERVIEW. PART 1

Marina A. Nikitina\*, Irina M. Chernukha

V. M. Gorbatov Federal Research Center for Food Systems of Russian Academy of Sciences, Moscow, Russia

**Keywords:** nonparametric statistics, Rosenbaum Q-test, Mann-Whitney U-test, sign G-test, Wilcoxon T-test

## Abstract

Daily, researcher faces the need to compare two or more observation groups obtained under different conditions in order to confirm or argue against a scientific hypothesis. At this stage, it is necessary to choose the right method for statistical analysis. If the statistical prerequisites are not met, it is advisable to choose nonparametric analysis. Statistical analysis consists of two stages: estimating model parameters and testing statistical hypotheses. After that, the interpretation of the mathematical processing results in the context of the research object is mandatory. The article provides an overview of two groups of nonparametric tests: 1) to identify differences in indicator distribution; 2) to assess shift reliability in the values of the studied indicator. The first group includes: 1) Rosenbaum Q-test, which is used to assess the differences by the level of any quantified indicator between two unrelated samplings; 2) Mann-Whitney U-test, which is required to test the statistical homogeneity hypothesis of two unrelated samplings, i. e. to assess the differences by the level of any quantified indicator between two samplings. The second group includes sign G-test and Wilcoxon T-test intended to determine the shift reliability of the related samplings, for example, when measuring the indicator in the same group of subjects before and after some exposure. Examples are given; step-by-step application of each test is described. The first part of the article describes simple nonparametric methods. The second part describes nonparametric tests for testing hypotheses of distribution type (Pearson's chi-squared test, Kolmogorov test) and nonparametric tests for testing hypotheses of sampling homogeneity (Pearson's chi-squared test for testing sampling homogeneity, Kolmogorov-Smirnov test).

## Funding:

The research was supported by state assignment of V. M. Gorbatov Federal Research Centre for Food Systems of RAS, scientific research No. FNEN-2019-0008.

## Introduction

The starting point for state-of-the-art statistical methods development is 1900. In 1901, the English mathematician, statistician, biologist and philosopher Karl Pearson, together with Francis Galton and Walter Weldon, founded the *Biometrika* journal (<https://academic.oup.com/biomet>) to promote the introduction of mathematical methods in biology. In 1925, Karl Pearson created another journal, *Annals of Human Genetics* (<https://onlinelibrary.wiley.com/journal/14691809>) dedicated to human genetics.

First third of the twentieth century passed under the sign of parametric statistics, which is defined as a section of statistics "assuming that the sampling belongs to the general population, which may be sufficiently accurately and adequately modeled by a probability distribution with specified set of parameters" [1]. Its main object is sampling from distributions described by one or several parameters. Methods based on the data analysis from parametric distributions described by the Pearson curves were studied. The normal distribution (Gaussian distribution) was the most popular one. Exponential and logarithmic normal distributions, Weibull-Gnedenko distributions, gamma distributions, binomial and hypergeometric distributions, Poisson distributions, etc. were used. Pearson test, Student test, and Fisher test were used to test hypotheses. The method of maximum likelihood and analysis of variance

were proposed, and the main ideas for experiment design were stated [1-7].

Simultaneously with parametric statistics, in the works of Charles Edward Spearman and Maurice George Kendall, the first nonparametric methods based on rank correlation coefficients were introduced. Now the above methods are named after these statisticians. But nonparametric statistics became a noticeable part of statistics only since the second third of the twentieth century. Originally, nonparametric methods were designed to test statistical hypotheses (about one-dimensional probability distributions). The most famous nonparametric tests are Kolmogorov-Smirnov tests developed in 1930s, Wilcoxon rank tests and the Mann-Whitney U-test developed in 1940s-1950s, and, of course, rank correlation coefficients (1904-1930) by C. Spearman and M. Kendall [8-24].

Nonparametric tests are tests that do not include distribution parameters in the calculation formula and do not suppose knowledge of the distribution function. These tests use frequencies or ranks. They are used when comparing samplings with nominal and serial indicators, as well as for samplings measured on quantitative scales for which the distribution law is unknown or differs from normal one.

A feature of the nonparametric methods is that the probability distribution is considered completely unknown, and the problems are stated only in terms of dif-

ferences between classes or within classes of unknown distributions.

If an alternative hypothesis contains a statement about the expected direction of differences, then such a hypothesis is called “directional”, otherwise the hypothesis is called “non-directional”.

The level of statistical significance or critical values of tests are determined differently when testing directional and non-directional hypotheses.

When testing the directional hypothesis, one-sided test is used, while two-sided test is used for the non-directional hypothesis. The two-sided test is more stringent because it tests for differences in both directions, and therefore the empirical value of test that previously met a significance level of  $\leq 0.05$  now corresponds to a level of  $\leq 0.1$ .

Many problems in research and practical work are associated with the need to compare two or more groups of observations obtained under different conditions, as well as the need to classify observations (rank them to one of several classes). The solution for such problems is possible only when it is determined what is the difference (or identity) for the groups of observations. When the compared sets of numbers are so different that they do not overlap, the researcher has no doubt that the sets are different. Difficulties begin when the sets overlap.

In the first part of the article, we will discuss two groups of nonparametric tests: 1) to identify differences in indicator distribution; 2) to estimate shift reliability in the values of the studied indicator.

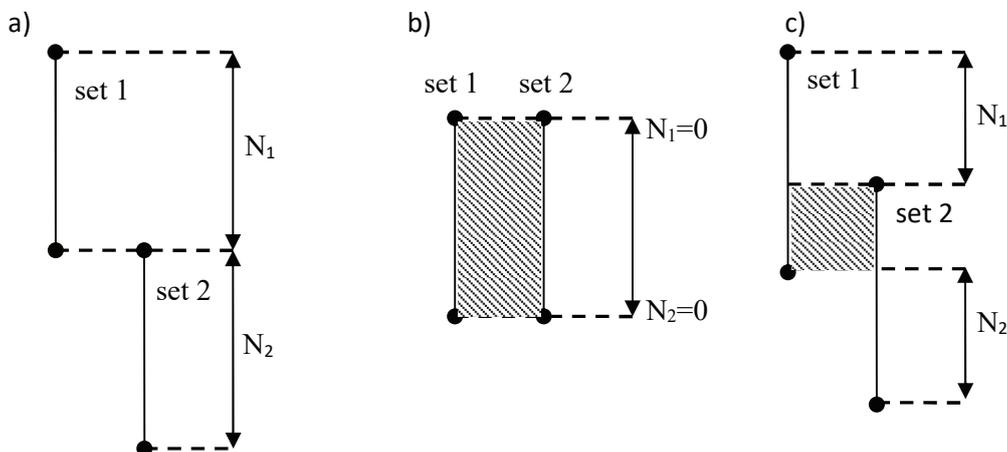
The aim of the research was to determine the application limits of the tests and methods in nonparametric statistics.

**Nonparametric tests to identify differences in indicator distribution**

This group of tests includes Rosenbaum Q-test [21], Mann-Whitney U-test [20].

**1. Rosenbaum Q-test**

The test is used to assess the differences by the level of any quantified indicator between two unrelated samplings [21].



**Figure 1.** Possible ratios for sets of values in two samplings.  $N_1$  is the area of values in the 1st set, which are higher than the maximum value in the 2nd set;  $N_2$  is the area of values in the 2nd set, which are less than the minimum value in the 1st set; overlapping areas of two sets are marked with hatching

Figure 1 shows a possible arrangement of two sets (samplings). Two compared sets are ideally arranged if they may be represented as in Figure 1 (a):  $N_1$  is the number of observations in one sampling exceeding the maximum value in the other sampling;  $N_2$  is the number of observations of one sampling, the values of which are less than the minimum value in the other sampling.

In case (a), there is no overlap between the values of both sets. Therefore, there is a statistically significant difference between the two sets (i. e., the difference is significant). In Figure 1b, both sets are at the same level. Therefore, the difference is not significant. If there are equal values in the compared sets, they should be placed exactly opposite each other. In Figure 1c, sets partially overlap, but the first set is higher than the second one.

Let’s explain why  $N_1=0$  and  $N_2=0$  in Figure 1b.

$N_1$  are numbers from sampling 1 that are higher than the maximum value in sampling 2. In Figure 1b, there are no such numbers in sampling 2.  $N_2$  are numbers from sampling 2 that are less than the minimum value in sampling 1. In Figure 1b, there are no such numbers in sampling 1. Thus,  $N_1$  and  $N_2$  in Figure 1b are equal to 0.

The Q-value (Rosenbaum Q-test) is equal to the sum of  $N_1$  and  $N_2$ . The higher it is, the differences are more reliable, i. e.  $\sum_{i=1}^2 N_i$  (Figure 1a)  $\geq \sum_{i=1}^2 N_i$  (Figure 1c)  $\geq \sum_{i=1}^2 N_i$  (Figure 1b). Experimental value of Rosenbaum Q-test is calculated by the formula:

$$Q_{exp} = N_1 + N_2 \tag{1}$$

The following hypotheses are tested:

$H_0$ : Indicator level in sampling 1 does not exceed indicator level in sampling 2.

$H_1$ : Indicator level in sampling 1 exceeds indicator level in sampling 2.

**Application of Rosenbaum Q-test**

1. Arrange the values of indicator in both samplings in descending (or ascending) order. Consider sampling 1 the one that is supposed to include higher values, and sampling 2 the one that is supposed to include lower values (Figure 1c).

2. Determine the highest (maximum) value in sampling 2.
3. Count the number of values in sampling 1 that are higher than the maximum value in sampling 2 ( $N_1$ ).
4. Determine the lowest (minimum) value in sampling 1.
5. Count the number of values in sampling 2 that are lower than the minimum value in sampling 1 ( $N_2$ ).
6. Calculate  $Q_{exp}$  by the formula (1).
7. In the table of critical values presented in [7, 21, 25, 26] for given  $n_1$  and  $n_2$  and the significance level  $\alpha$ , find the critical value  $Q_{cr}$ .
8. If  $Q_{exp}$  is equal to  $Q_{cr}$  or exceeds it,  $H_0$  is rejected.

**Limitations of Rosenbaum Q-test**

1. Measurement may be carried out on a scale of order, intervals and ratios.
2. Samplings must be unrelated.
3. Each sampling must contain at least 11 observations. Furthermore, the volumes of samplings should be approximately the same:
  - a) if both samplings contain less than 50 observations, then the difference between the volumes of samplings should be not more than 10 observations;
  - b) if each sampling contains more than 50 observations, but less than 100 observations, then the difference between the volumes of samplings should be not more than 20 observations;
  - c) if each sampling contains more than 100 observations, then it is allowed that one of the samplings is 1.5 to 2 times larger than the other one.
4. If the numbers of observations  $n_1$  and  $n_2$  are  $\geq 26$ , the following values may be used:

$$Q_{cr} = \begin{cases} 8 & \text{at } \alpha \leq 0.05 \\ 10 & \text{at } \alpha \leq 0.01 \end{cases}$$

5. The ranges of value distribution for the two samplings should not overlap. In this case, the Q-test is not applicable. Therefore, Mann-Whitney U-test should be used.

**Example**

*Given data:* Heart rate was measured in two groups of animals. One group (11 animals) includes animals before the experiment, the other group (12 animals) includes animals after the experiment. The results are presented in Table 1. Is it possible to state that one of the groups is superior to the other one in terms of heart rate at the significance level? [25].

*Solution.* All the Q-test limitations are met, so Rosenbaum Q-test may be used to identify differences in indicator distribution.

Let's arrange the values in both samplings in ascending order, and then state the hypotheses:

$H_0$ : The first group of animals does not exceed the second group of animals in heart rate.

$H_1$ : The first group of animals exceeds the second group of animals in heart rate.

The calculations are presented in Table 2. The first set is the one that is "higher" — set No.  $N_1$ , and the second set is the one that is "lower" — set No.  $N_2$ . The maximum value

in the second sampling is 122 and the minimum value in the first sampling is 92.

**Table 1. Given data for the problem**

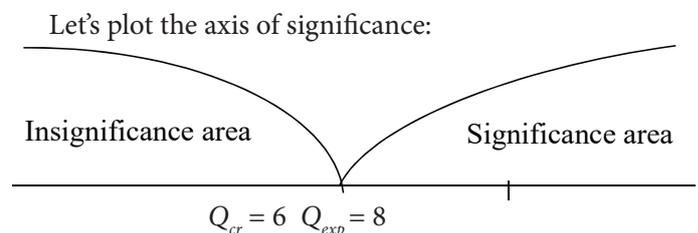
1st group of animals		2nd group of animals	
No.	Heart rate	No.	Heart rate
1	92	1	77
2	98	2	81
3	104	3	81
4	107	4	84
5	107	5	89
6	122	6	95
7	122	7	101
8	122	8	103
9	127	9	107
10	133	10	110
11	137	11	114
		12	122

From Table 2, we determine the number of values in the first set (137, 133, 127) that are higher than the maximum value in the second set ( $N_1 = 3$ ) and the number of values in the second set (89, 84, 81, 77) that are lower than the minimum value in the first set ( $N_2 = 5$ ).

We calculate  $Q_{exp}$  according to the formula (1):  $Q_{exp} = N_1 + N_2 = 3 + 5 = 8$ . From the table of critical values presented in [7, 21, 25, 26], we can find critical values for the given sampling volume and the significance level. At  $n_1 = 11$ ,  $n_2 = 12$  and we have  $Q_{cr} = 6$ .

**Table 2. Calculation results**

1st group of animals		2nd group of animals	
No.	Heart rate	No.	Heart rate
1	137		
2	133		
3	127		
-----			
4	122	1	122
5	122		
6	122		
		2	114
		3	110
7	107	4	107
8	107		
9	104		
		5	103
		6	101
10	98		
		7	95
11	92		
-----			
		8	89
		9	84
		10	81
		11	81
		12	77



Since  $Q_{exp} > Q_{cr}$ ,  $H_0$  is rejected. And is accepted. Thus, the second group of animals is superior to the animals of the first group in heart rate (at  $\alpha < 0.05$ ).

In [27], the authors applied the method of mathematical statistics (Rosenbaum Q-test) when finding the optimal technology for developing professional competencies among the target group of students studying business informatics. In [28], the exercise training effect on the health and condition in two groups of women of different ages was assessed by Student t-test (t) and Rosenbaum Q-test (Q). In [29], to compare and evaluate the data in five groups of 110 students, in a study on how English can be effectively implemented in the training of students of technical specialties for developing professional skills, the statistical method with Rosenbaum Q-test was used.

2. Mann-Whitney U-test

The test is designed to test the statistical homogeneity hypothesis of two unrelated samplings, i. e. to assess the differences by the level of any quantified indicator between two samplings [20]. It allows to differentiate between small samplings when  $n_1, n_2 \geq 3$  or  $n_1 = 2, n_2 \geq 5$  ( $n_1$  and  $n_2$  are volumes of samplings), and is more powerful than the Rosenbaum Q-test. The Rosenbaum Q-test is designed for low-volume samplings. Thus, Rosenbaum Q-test is a particular case of Whitney-Mann test.

Two samplings,  $x_1, \dots, x_n$  (sampling X) and  $y_1, \dots, y_n$  (sampling Y) with volumes of  $n_1$  and  $n_2$  are studied. Let's denote the distribution law of the first sampling by  $F$ , and the distribution law of the second sampling by  $G$ . The laws of distributions  $F$  and  $G$  are continuous. From here it follows with probability 1, that among the numbers  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$ , there are no overlapped ones.

Let's test the null hypothesis  $H_0: F = G$ .

All possibilities of  $F \neq G$  may be the alternative hypotheses. However, the test is not able to detect all possible deviations from  $H_0$ . First of all, this test is intended to test the  $H_0$  hypothesis against the alternative of  $F \geq G$  (right-sided alternative, "overflow" of probabilities to the right) or the alternative of  $F \leq G$  (left-sided alternative, "shift" of probabilities to the left). Combining both possibilities (two-sided alternative) may also be studied.

The Mann-Whitney U-test is based on a pairwise comparison of the results from the first and the second

sampling. For values of  $x_i < y_j$ , an event means "success", and any event with  $x_i > y_j$  means "failure". By changing  $i$  from 1 to  $n_1$  and  $j$  from 1 to  $n_2$ , we have  $n_1 \times n_2$  pairwise comparisons of values in samplings X and Y.

Let's denote the number of successes in these pairwise comparisons by  $U$ .  $U$ -value can take any integer value from 0 to  $n_1 \times n_2$ . The random  $U$ -value is called the Mann-Whitney statistics.

This test is particularly effective when testing an alternative hypothesis of dominance, and its particular case, i. e. the hypothesis of the right shift, when the values in the first sampling are higher than the values in the second sampling.

The test determines if the overlapping values area between the two sets is small enough. The first set, sampling or group is the set of values in which the values, according to preliminary estimates, are higher. And the 2nd set is the one where the values are supposed to be lower. The smaller the area of overlapping values, the more likely the differences are significant. These differences are sometimes referred to as differences in the location of the two samplings. The experimental value  $U_{exp}$  reflects how large is the overlapping area between sets. Therefore, the smaller  $U_{exp}$ , the more likely the differences are significant. Figure 2 shows three of the many possible relationships between two sets of values.

In the first variant (Figure 2a), the second set is lower than the first one, and the sets almost do not overlap. This means that the values are almost different. The overlapping area is too small to hide the differences between sets. The differences between seem to be significant. In the second variant (Figure 2b), the second set is also lower than the first one, but the area of overlapping values between the two sets is quite large. It may not yet reach a critical value, when the differences have to be recognized as insignificant. Finally, in the third variant (Figure 2c), the second set is lower than the first one, but the overlapping area is so large that the differences between the sets are insignificant.

The following hypotheses are tested:

$H_0$ : indicator level in group 2 is not lower than indicator level in group 1.

$H_1$ : indicator level in group 2 is lower than indicator level in group 1.

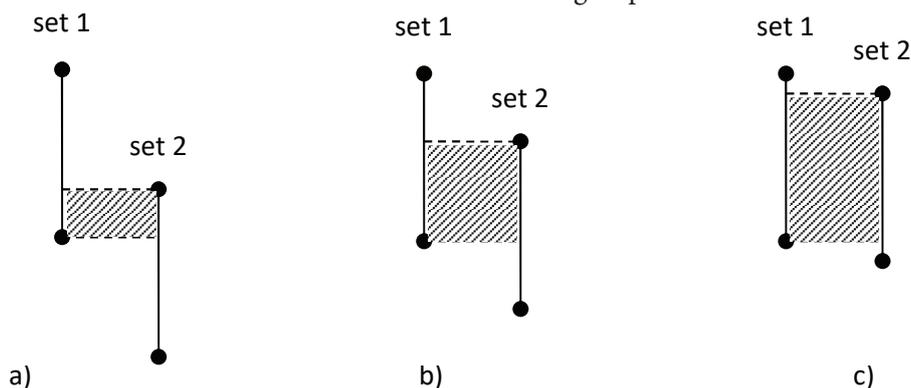


Figure 2. Possible ratios for sets of values in two samplings; overlapping areas are marked with hatching

**Application of Mann-Whitney U-test**

1. Pool the values for both groups in the order of indicator rising, regardless of what sampling they belong to.
2. Rank the values by assigning a lower rank to a lower value. Total ranks ( $n_1 + n_2$ ).
3. Calculate the sum of the ranks for the values in the first sampling and the sum of the ranks for the values in the second sampling. Determine the larger of the two sums of the ranks.
4. Determine  $U_{exp}$  value by the formula:

$$U_{exp} = n_1 \cdot n_2 + \frac{n_x \cdot (n_x + 1)}{2} - T_x \quad (2)$$

where  $n_1, n_2$  are volumes of samplings 1 and 2;  $T_x$  is the larger of the two rank sums;  $n_x$  is the volume of sampling with the larger sum of the ranks.

5. Determine the critical values  $U_{cr}$  for the given  $n_1, n_2$  and the significance level  $\alpha$ . If  $U_{exp} > U_{cr}$ , then  $H_0$  is accepted. If  $U_{exp} < U_{cr}$ , then  $H_0$  is rejected. The lower the U-value, the higher the reliability of the differences.

**Limitations of Mann-Whitney U-test**

1. Each sampling must contain at least 3 observations:  $n_1, n_2 \geq 3$ ; it is allowed that one of the sampling includes 2 observations, but then the second one should include at least 5 observations.
2. Each sampling should contain no more than 60 observations:  $n_1, n_2 \leq 60$ .

**Example**

*Given data.* In the control group (without treatment) and in the experimental group (with treatment), the death of animals was established in time intervals (minutes) after the introduction of a toxic substance [7]. The results are shown in Table 3.

**Table 3. Given data for the problem**

Sampling No.	1	2	3	4	5	6	7	8	9
No. 1 — without treatment	39	37	45	5	27	27	31		
No. 2 — with treatment	45	9	67	42	42	31	34	49	99

Let's state the hypotheses:

$H_0$ : the difference between the indicators in the control group and the experimental group is statistically insignificant.

$H_1$ : the difference between the indicators in the control group and the experimental group is statistically significant.

*Solution.* In this case, the Rosenbaum Q-test is not applicable, since the volume of both samplings is less than 11. Let's apply the Mann-Whitney U-test.

Let's pool both samplings arranging their values in ascending order, while indicating which sampling they belong to and ranking them. The data are summarized in Table 4.

Rank is a serial number. The lowest value has rank 1.

The value of 5 — rank 1.

The value of 9 — rank 2.

We have the value of 27 twice, so it takes positions (ranks) 3 and 4. We add (3 and 4) and divide by 2 (since 27 occurs twice in sampling). Thus  $(3 + 4)/2 = 3.5$ . The rank of 27 is 3.5.

The value of 31 occurs twice in sampling taking positions 5 and 6. Similarly, we add 5 and 6 and divide by 2, so we get rank 5.5, etc.

**Table 4. Pooled data from two samplings**

Sampling No.	Values	rank	Sampling No.	Values	rank
1	5	1	1	39	9
2	9	2	2	42	10.5
1	27	3.5	2	42	10.5
1	27	3.5	1	45	12.5
1	31	5.5	2	45	12.5
2	31	5.5	2	49	14
2	34	7	2	67	15
1	37	8	2	99	16

Let's check the correctness of the ranking:

$$\sum(R_i) = 1 + 2 + 3.5 + 3.5 + 5.5 + 5.5 + 7 + \dots + 14 + 15 + 16 = 136$$

Using the formula that calculates the sum of ranks in sampling, we determine the correctness of the ranks. As a result:

$$\sum(R_i) = \frac{N \cdot (N + 1)}{2} = \frac{16 \cdot (16 + 1)}{2} = 136$$

Thus, the ranks are assigned correctly.

Calculating the ranks in the first and the second sampling:

$$R_1 = 1 + 3.5 + 3.5 + 5.5 + 8 + 9 + 12.5 = 43$$

$$R_2 = 2 + 5.5 + 7 + 10.5 + 10.5 + 12.5 + 14 + 15 + 16 = 93$$

Checking:  $43 + 93 = 136$ .

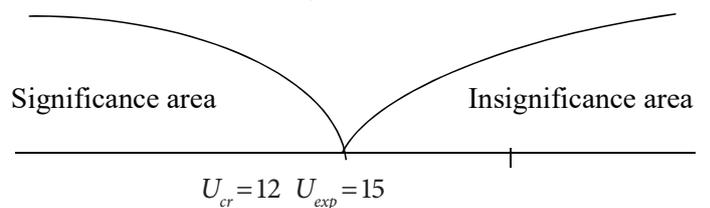
The larger of the two rank sums  $T_x = 93$  (corresponds to the second sampling  $n_2 = 9$ ), i. e.  $n_x = 9$ .

Let's find the experimental value of test by the formula:

$$U_{exp} = n_1 \cdot n_2 + \frac{n_x \cdot (n_x + 1)}{2} - T_x = 7 \cdot 9 + \frac{9 \cdot (9 + 1)}{2} - 93 = 15$$

We can find the critical value of test according to the table presented in [7, 21, 25, 26]. In this case, we are testing the non-directional alternative hypothesis. With  $n_1 = 7$  and  $n_2 = 9$ ,  $U_{cr} = 12$  for  $\alpha \leq 0.05$ .

Let's plot the axis of significance:



Since  $U_{exp} > U_{cr}$ ,  $H_0$  is accepted. Thus, the significant differences between the indicators of the control and the experimental group has not been established.

In [30], the protein damage of individual protease inhibitors of the serpin superfamily in CRSwNP was studied. The damage of protease inhibitors by fibrinolysis in tissues was compared with exosome samples to assess the potential of “liquid biopsy” non-invasive exosomes for CRSwNP. Semi-quantitative analysis of Western blots was performed using the Mann-Whitney U-test. Effectiveness comparison of the standard simulation training in emergency care with a more comprehensive simulation program for emergency care is shown by the authors in [31]. The one-sided t-test and the Mann-Whitney U-test were used to analyze the Objective Structured Clinical Examination (OSCE) scores. The assessment of the effect of ketorolac intravenous injections on pain relief before and after using a tourniquet in orthopedic surgery on the femur and tibia was shown in [32]. This double-blind clinical study enrolled one hundred patients aged 15 to 75, candidates for hip and lower leg surgery. Visual pain tests during recovery were measured and analyzed using descriptive (mean and percentage) and logical statistics (Mann-Whitney U-test). The study [33] assessed the implementation of patient-centered interventions. 34 patients were examined with elective total knee arthroplasty. Endpoints were duration of induction period (primary), duration of hospital stay, pain at rest and pain at physical activity on postoperative day 1 (numeric analogue scale), and surgical release progress (MBF) on postoperative days 1, 3, and 6 (secondary). Group inferiority comparisons were made using the Wilcoxon T-test and Mann-Whitney U-test. Work [34] studied MRI of sodium ( $^{23}\text{Na}$ ) at a field strength of 7 Tesla (T) as a biomarker of tumor stage, isocitrate dehydrogenase (IDH) mutation, and O6-methylguanine-DNA methyltransferase (MGMT) promoter methylation in patients with glioma. The difference in median  $^{23}\text{Na}$  concentration throughout the tumor area was compared between gliomas with IDH mutation and wild-type IDH, as well as between glioblastomas with methylated and non-methylated MGMT using the Mann-Whitney U-test.

#### *Nonparametric tests for assessing shift reliability in the values of the studied indicator*

This group of tests includes sign G-test and Wilcoxon T-test.

Let a random variable have a distribution  $F(x)$ , and another random variable has a distribution  $S(x)$ , and  $S(x) = F(x - \Delta)$ , where  $\Delta$  is an abstract parameter characterizing the state of the object under study.

Two samplings,  $x_1, \dots, x_n$  (sampling from  $F(x)$ ) and  $y_1, \dots, y_n$  (sampling from  $S(x)$ ), with volumes  $n$  are studied. Furthermore,  $F(x)$  and  $S(x)$  are unknown distribution functions.

It is required to test the null hypothesis  $H_0: F(x) = S(x)$ , i. e.  $\Delta = 0$ . Three types of alternative hypotheses may be stated:

Directional hypotheses:

- $H_1^1: \Delta > 0;$
- $H_1^2: \Delta < 0;$

non-directional hypothesis:

- $H_1^3: \Delta \neq 0.$

The following hypotheses are tested:

$H_0$ : indicator level in group 2 is not lower than indicator level in group 1.

$H_1$ : indicator level in group 2 is lower than indicator level in group 1.

Thus, it is known that if the impact changes the indicator distribution, then the impact shifts it in a quite definite way, i. e. by the value of  $\Delta$ . If the shifts are compared in two unrelated samplings, for example, in the experimental and the control groups of subjects, then the Rosenbaum Q-test and the Mann-Whitney U-test should be used. To determine the shift reliability of related samplings, for example, when measuring indicator in the same group of subjects before and after some exposure, the sign G-test and the Wilcoxon T-test should be used.

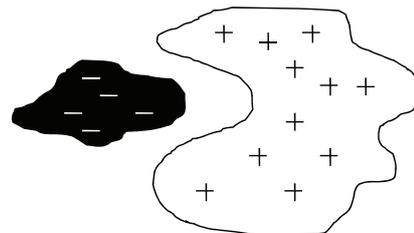
#### *1. Sign G-test*

Sign G-test is intended to establish a general shift direction of the studied indicator. It allows to determine, in which direction the values of the indicator in sampling generally change during the transition from the first measuring to the second one, i. e. whether the indicators change in the direction of improvement/increase or, conversely, in the direction of deterioration/decrease. Shift is the difference in the values of indicators measured after a certain impact and before it [35, 36].

The sign G-test is applicable both to those shifts that can only be determined qualitatively (for example, a change from a negative attitude to a positive one) and to those shifts that can be quantified.

Shifts that seem to be predominant are usual, and shifts in a rarer direction are unusual. If the indicators do not increase or decrease, then such shifts are called zero shifts and they are excluded from the sign G-test. Furthermore, the number of compared pairs is reduced by the number of such zero shifts.

The sign G-test determines whether there are too many unusual shifts observed to consider shift in usual direction to be predominant? The less unusual shifts, the more likely the prevalence of usual shift is reliable. Figure 3 shows usual shifts as a light area, and unusual shifts as a dark area.



**Figure 3.** A graphic representation of positive and negative shifts in the form of different areas: light area is positive shifts; dark area is negative shifts [26]

The following hypotheses are tested:

- $H_0$ : predominance of usual shift direction is random;
- $H_1$ : predominance of usual shift direction is not random.

### Limitations of sign G-test

The number of observations in both measurements should be not less than 5 and not more than 300.

### Application of sign G-test

1. Check if the limitations are met.
2. Enter the measurement data in the table (Table 5):

Table 5. The filled table

No.	Pre-exposure values	Post-exposure values	«After to before» shift
1			
2			
...			
n			

3. Shift is not quantified, but the difference sign “+” or “-” is added. If the difference is zero, then zero is written in the table.
4. Calculate the number of zero responses  $n_0$  and exclude them from consideration. As a result, the volume of sampling will decrease by the number of zero responses:  $n = n - n_0$ .
5. Determine the predominant direction of change. Shifts in the predominant direction are “usual”.
6. Determine the number of unusual shifts. Consider this number as the experimental value of G-test ( $G_{exp}$ ).
7. Determine the critical value  $G_{cr}$  for a given n and a specified significance level  $\alpha$ .
8. Compare  $G_{exp}$  and  $G_{cr}$ . If  $G_{exp} \leq G_{cr}$ ,  $H_0$  is rejected, i. e. shift in the typical direction may be considered reliable.

### Example

*Given data:* A group of 10 students completed a mathematics test. After checking the examination work, this group was asked to study an electronic manual on the same topic, after which a repeated test was carried out, which was assessed on a five-point scale. Has the knowledge of the students improved after studying the manual? [26].

*Solution.* The results of double examination work represent measurements in a serial scale. The volume of sampling is higher than 5, so the use of the sign G-test is possible. The results are summarized in Table 6:

Table 6. Given data for the problem

No.	First testing	Second testing	Sign of the difference
1	3	4	+
2	2	3	+
3	2	2	0
4	4	3	-
5	3	2	-
6	3	4	+
7	4	4	+
8	2	3	+
9	2	2	0
10	3	4	+

The number of zero responses  $n_0 = 2$ . Let's exclude them from consideration, therefore, volume of sampling  $n = 10 - 2 = 8$ .

The number of positive shifts is 6, and the number of negative shifts is 2. This means that the predominant direction is the positive one, therefore, shifts in this direction are usual, and shifts in the negative direction are unusual.

Let's state the hypotheses:

$H_0$ : predominance of usual shift direction is random, i. e. students' knowledge did not improve after studying the manual.

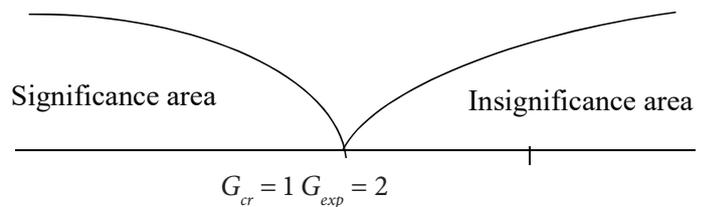
$H_1$ : predominance of usual shift direction is not random, i. e. students' knowledge improved after studying the manual (shift in the positive direction is reliable).

Let's test at the significance level  $\alpha \leq 0.05$ .

The number of unusual shifts is 2, i. e. (according to item 6).

For  $n = 8$  and  $\alpha \leq 0.05$  we can find from the table of critical values [7, 21, 25, 26].

Let's plot the axis of significance:



Since  $G_{exp} > G_{cr}$ , there is no reason to reject  $H_0$ , and the predominance of usual shift direction is random, so the students' knowledge did not improve after studying the manual.

### 2. Wilcoxon T-test

The test is used to compare values measured under two different conditions in the same sampling. It allows to establish not only the direction of changes, but also their intensity. It may be used to determine if the shift of indicators in one direction is more intense than in the other direction. Test is applicable in cases where indicators are measured at least on a serial scale, and the shifts between the second and first measurements can also be ordered. Therefore, they must vary within a considerably wide range [22, 23, 24].

The method compares the shift intensity in one direction with the shift intensity in another direction in absolute value. To do this, you need to rank all absolute shifts, and then summarize the ranks of shifts separately in the positive and negative directions. If shifts occur randomly, then the rank sums of their absolute values will be approximately equal. If the shift intensity in one of the directions outweighs, then the rank sum of absolute shifts in the opposite direction will be significantly lower than it would be with random changes.

In Figure 4, shifts in opposite directions are represented as different areas.

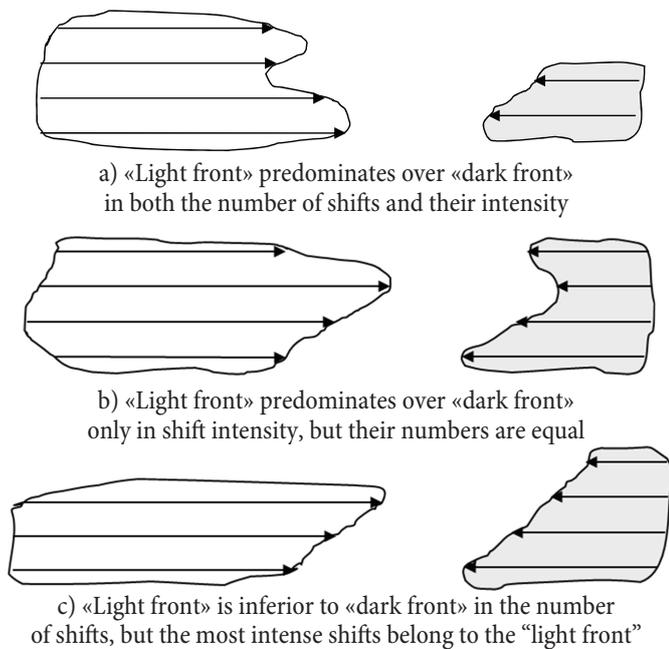


Figure 4. Types of the “light front” to “dark front” ratios; shifts in two different directions [26]

Initially, it is assumed that usual shift will shift in a more common direction, and unusual shift will shift in a rarer direction.

Let’s state the hypotheses:

$H_0$ : predominance of shifts between the initial and final indicators in one of the directions is unreliable (does not differ significantly from zero).

Alternative hypotheses:

Non-directional hypothesis:

$H_1^1$ : predominance of shifts between the initial and final indicators in one of the directions is reliable (significantly differs from zero).

Directional hypothesis:

$H_1^2$ : shift intensity in usual direction exceeds shift intensity in unusual direction (significantly differs from zero).

**Application of Wilcoxon T-test**

1. Make a list of pairs in any order.
2. Calculate the difference between the individual values in the second and the first measurements (before and after). Determine what will be considered the usual shift and state the corresponding hypotheses.
3. Convert the differences to absolute values and write them down in a separate column.
4. Rank the absolute values of the differences, assigning a lower rank to a lower value. Check the agreement of the received sum of ranks with the calculated one.
5. Note the ranks corresponding to shifts in unusual direction.
6. Calculate the sum of these ranks using the formula; where is rank of shifts with a rarer sign.
7. Determine the critical values  $T_{cr}$  for a given volume of sampling  $n$  and the significance level  $\alpha$ . If  $T_{exp} \leq T_{cr}$ , the null hypothesis is rejected, i. e. shift to the “usual” side in terms of intensity reliably predominates.

**Limitations of Wilcoxon T-test**

The minimum volume of sampling is 5.

Zero shifts are excluded and the number of observations is reduced by the number of these zero shifts. This limitation may not to be used by stating hypotheses that include no change: “Shift towards increasing values exceeds shift towards decreasing values and the tendency to keep them at the same level.”

**Example**

*Given data.* A pharmaceutical company researches a new drug. A group of 10 volunteers suffering from disease was selected. They had their body temperature measured before and 30 minutes after taking the new drug. The data are presented in Table 7. It is required to conclude about the significance of a decrease in body temperature as a result of the drug administration [7].

Table 7. Given data for the problem

n	Initial indicators (body temperature before drug administration)	Final indicators (body temperature after drug administration)	Difference of indicators (d)	Absolute difference	Rank of the difference
1	39	37.6	-1.4	1.4	7
2	39.5	38.7	-0.8	0.8	5
3	38.6	38.7	0.1	0.1	<b>1.5*</b>
4	39.1	38.5	-0.6	0.6	4
5	40.1	38.6	-1.5	1.5	8
6	39.3	37.5	-1.8	1.8	9
7	38.9	38.8	-0.1	0.1	1.5
8	39.2	38	-1.2	1.2	6
9	39.8	39.8	0	0	0
10	38.8	39.3	0.5	0.5	<b>3*</b>

\* Unusual ranks are in bold italics. In this case, the atypical situation is that the temperature should decrease, but it increases.

*Solution.* Let’s state the hypotheses:

$H_0$ : predominance of shifts between the initial and final indicators in one of the directions is unreliable (does not differ significantly from zero).

$H_1$ : predominance of shifts between the initial and final indicators in one of the directions is reliable (significantly differs from zero).

Let’s test the null hypothesis at the significance level  $\alpha \leq 0.05$ .

The sum of unusual ranks is equal to the experimental value of the Wilcoxon T-test:

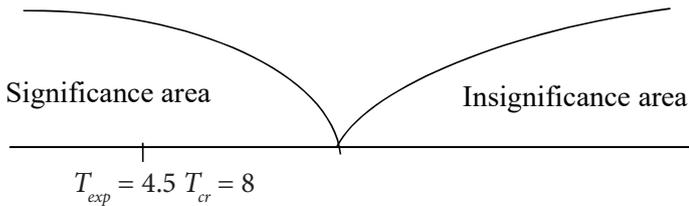
$$T_{exp} = 1.5 + 3 = 4.5.$$

Usual shift of the indicator (body temperature) is its decrease.

From the table of critical values presented in [7, 21, 25, 26], we can find.

Since the direction of the differences was not predicted in advance, we test the hypothesis for the two-sided test. For  $n = 9$  (zero shift is excluded) and  $\alpha = 0.05$   $T_{cr} = 8$ .

Let's plot the axis of significance:



Since  $T_{cr} > T_{exp}$ , the null hypothesis is rejected. It can be stated that the reliability between the initial and final indicators in one of the directions is established.

### Example

Given data:

In a group of students, training was conducted on the development of creative thinking. Tests were carried out before and after the training (stimulating effect should increase creativity, i. e. increase the score). The data are presented in Table 8. Is the hypothesis from the training providers confirmed that the training contributes to the development of creative thinking at the significance level  $\alpha \leq 0.05$ ?

Table 8. Given data for the problem and calculation results

n	Before training	After training	Difference of indicators (d)	Absolute difference	Rank of the difference
1	19	17	-2	2	4.5
2	20	26	6	6	11
3	18	20	2	2	4.5
4	15	18	3	3	6.5
5	29	30	1	1	2
6	21	25	4	4	8.5
7	21	28	7	7	12
8	18	19	1	1	2
9	21	20	-1	1	2
10	23	27	4	4	8.5
11	14	19	5	5	10
12	10	13	3	3	6.5

*Solution.* The first step in calculating the T-test is to subtract each individual "before" value from the "after" value. The calculation results in the form of the difference of indicators (d) are presented in Table 8. Two of the obtained differences are negative, and 10 of the differences are positive. Therefore, the usual direction is positive.

Let's state the hypotheses:

$H_0$ : shift intensity in usual direction does not exceed shift intensity in unusual direction.

$H_1$ : shift intensity in usual direction exceeds shift intensity in unusual direction.

In the next step, all shifts, regardless of their sign, must be ranked.

We check the correctness of the ranking.

$$\sum(R_i) = 2 + 2 + 2 + 4.5 + 6.5 + 6.5 + 8.5 + 8.5 + 10 + 11 + 12 = 78$$

We check by the formula:

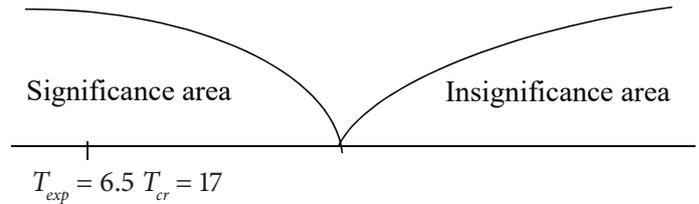
$$\sum(R_i) = \frac{N \cdot (N+1)}{2} = \frac{12 \cdot (12+1)}{2} = 78$$

Consequently, ranks are assigned correctly.

Note those shifts that are unusual, i. e. negative ones (in Table 8 they are marked in bold italics). The sum of the ranks of these unusual shifts is the experimental value of the T-test:  $T_{exp} = 2 + 4.5 = 6.5$ .

In this problem, the alternative hypothesis is directional because the direction of the differences is predicted in advance, so one-sided test should be used to test hypotheses. For  $n = 12$  and  $\alpha \leq 0.05$ , from the table of critical values presented in [7, 21, 25, 26], we can find  $T_{cr} = 17$ .

Let's plot the axis of significance:



Since  $T_{cr} \geq T_{exp}$ , the null hypothesis is rejected. It can be stated that shift intensity in usual direction exceeds shift intensity in unusual direction, therefore, the initial assumption about the effectiveness of the stimulating effect is confirmed.

Let the following test scores to be obtained for another group of subjects (Table 9). 6 differences are negative, and 4 differences are positive, i. e. it is impossible to state a hypothesis corresponding to the initial assumption. It is possible to state a hypothesis assuming the insignificance of the indicator shift in the direction of decreasing:

$H_0$ : shift intensity towards decreasing creativity does not exceed shift intensity towards its increase.

$H_1$ : shift intensity towards decreasing creativity exceeds shift intensity towards its increase.

Table 9. Given data for the problem and calculation results

n	Before training	After training	Difference of indicators (d)	Absolute difference	Rank of the difference
1	19	17	-2	2	3.5
2	23	20	-3	3	5.5
3	18	20	2	2	3.5
4	15	18	3	3	5.5
5	25	17	-8	8	10
6	21	15	-6	6	9
7	21	20	-1	1	1.5
8	18	19	1	1	1.5
9	24	20	-4	4	7
10	23	28	5	5	8

Let's rank the absolute differences and check the correctness of the ranking.

$$\sum(R_i) = 1.5 + 1.5 + 3.5 + 3.5 + 5.5 + 5.5 + 7 + 8 + 9 + 10 = 55$$

By the formula we obtain:

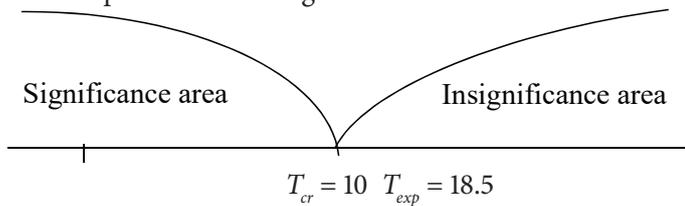
$$\sum(R_i) = \frac{N \cdot (N+1)}{2} = \frac{10 \cdot (10+1)}{2} = 55$$

Consequently, ranks are assigned correctly.

Note those shifts that are unusual, i. e. positive ones (in Table 9 they are marked in italics). Let's calculate the experimental value of T-test:  $T_{exp} = 3.5 + 5.5 + 1.5 + 8 = 18.5$ .

Since the direction of the differences is predicted in advance, a one-sided test is used to test the hypotheses. For  $n = 10$  and  $\alpha \leq 0.05$   $T_{cr} = 10$ . [7,21,25,26].

Let's plot the axis of significance



Since  $T_{cr} \geq T_{exp}$ , the null hypothesis is accepted. It can be stated that the intensity of the negative shift does not exceed the intensity of the positive shift.

The Wilcoxon T-test [37–49] is the most applicable for statistical processing in scientific researches. In [37], a study was conducted to validate the ecfX-targeted qPCR protocol designed to detect all viable *P. aeruginosa* bacteria and identify forms in sputum samples from patients with cystic fibrosis. The study involved 115 *P. aeruginosa* strains of various origins and 10 other strains in 88 cystic fibrosis sputum samples, 41 of which were positive and 47 were negative. The statistical significance of the results was assessed using the Wilcoxon T-test and Student t-test. The work [38] is aimed at improving diabetes care. The study is to assess the improvement of group care, which directly affects the quality of care for patients diagnosed with diabetes. Changes in laboratory values of A1c hemoglobin and LDL cholesterol were analyzed using the paired t-test and verified using the Wilcoxon T-test. The aim of work [39] was to conduct a study to provide additional information on the health benefits of gluten-free foods and to test the hypothesis that both self-selected and recommended diets are less nutritious when gluten-free diets are followed. Diet differences were assessed using paired t-tests and ranks of the Wilcoxon T-test. Assessment of the amino acid profile in trauma patients and assessment of the nutritional mode effect and the addition of exogenous ALA-GLN dipeptide on plasma amino acid concentration were conducted in [40]. Differences between plasma amino acid concentrations on day 1 and day 6 were assessed using Student t-test, Mann-Whitney U-test, and Wilcoxon T-test. In [41], a study of weight recovery after gastric bypass by increasing the consumption of high-calorie liquid and soft foods was carried out. Each patient with dysphagia was compared with 4 control patients based on the age, body mass index and time after surgery. Statistical analysis was performed using the Wilcoxon T-test. A study on the treatment of equine laminitis with and without acupuncture is reviewed in [42]. The study involved 14 horses with chronic laminitis. Lameness was statistically analyzed using the paired t-test and the sign ranked Wilcoxon T-test at  $p < 0.05$ . An analysis of the effect of yerba mate daily consumption on serum lipids and body composition in overweight women is shown in [43]. 119 overweight 25 to 50 years old women were randomized into three groups: mate and regime (MD),

mate without regime (M), water and regime (AD). For 12 weeks, the M and MD groups were supplemented with mate, while the AD and MD groups followed a hypocaloric regime. Anthropometric measurements and blood tests (total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides) were performed at the beginning and in the end of the study. Statistical analysis was performed using Student t-test and Wilcoxon T-test. Determination of the effect of a new phytosterol emulsion for nutritional supplements (1.5 g/day of phytosterol equivalent) on the concentration of low-density lipoprotein cholesterol was studied in [44]. Thirty-two healthy adults were randomized to receive placebo or treatment with a washout period followed by placebo or treatment. Each phase lasted one month. Secondary endpoints related to cardiovascular health were assessed. Study management, including screening, recruitment, monitoring, compliance and data collection, was carried out remotely (off-site clinical trial) using a new virtual instrument. The effect of phytosterol supplementation was assessed using Student t-test and Wilcoxon T-test. Changes in radiomic characteristics based on cone computed tomography (CBCT) during treatment and quantification of the potential relationship between CBCT-based delta radiomics characteristics and overall survival in locally advanced lung cancer patients have been reviewed [45]. 23 patients were identified and 658 radiomic characteristics were calculated for each of the 11 CBCT images in each patient. The choice of characteristics was made on the basis of repeatability, contour uncertainty resistance, and lack of redundancy. Evaluation of the predictive ability for survival was carried out using the t-test and Wilcoxon T-test. The work [46] investigated the effect of smoking on periodontal disease. Comparison of 441 daily smokers with 441 non-smokers was performed using the Wilcoxon T-test and t-test. The study [47] evaluated the reproducibility of 9 instruments for the analysis of differential expression in scRNA-seq data. Statistical analysis was performed using the t-test and Wilcoxon T-test. The study [48] was conducted from September 2018 to August 2019 and included 101 athletes with disabilities (13 Paralympic disciplines) from Brasilia, Federal District. Food intake was assessed based on two or four inconsistent nutritional reviews over a 24-hour period, in which para-athletes reported all food, drinks, and supplements consumed in the previous 24 hours. Diet analysis was performed using the method of the National Cancer Institute. Comparisons between fellows and athletes not receiving a fellowship were made using the Student t-test for parametric variables and Wilcoxon T-test for nonparametric variables. Research [49] was focused on the recruitment and retention of women in Type 2 Diabetes Women Health Program, and to assess the program's initial effectiveness in improving well-being. A 12-week wellness event enrolled 70 middle-aged women with type 2 diabetes. Analyzes were performed using chi-squared tests, McNemar test, paired t-test, and Wilcoxon T-test.

## Conclusion

The described nonparametric statistical tests are much less laborious than parametric ones, and for distributions far from normal (Gaussian), they are more efficient as well. It is completely insufficient for a current researcher to use one of the methods for statistical assessment of the differences between two groups of observations. In each case, a suitable test must be selected. This allows not only to in-

crease the efficiency of statistical processing and reduce its labor intensity, but also to get a clear idea of the comparison reliability of experimental results. The second part will review nonparametric tests for testing hypotheses of distribution type (Pearson's chi-squared test, Kolmogorov test) and nonparametric tests for testing hypotheses of sampling homogeneity (Pearson's chi-squared test for testing sampling homogeneity, Kolmogorov-Smirnov test).

## REFERENCES

1. Fisher, R.A. (1992). Statistical methods for research workers. Chapter in a book: Breakthroughs in statistics. Springer Series in Statistics (Perspectives in Statistics). New York: Springer. 1992. <https://doi.org/10.1007/978-1-4612-4380-96>
2. Plokhinskiy, N.A. (1978). Mathematical methods in biology. Moscow: Moscow State University. 1978. (In Russian)
3. Glants, S. (1998). Medical and biological statistics. Moscow: Practice. 1998. (In Russian)
4. Lakin, G.F. (1990). Biometrics. Moscow: High School. 1990. (in Russian)
5. Rokitskiy, P.F. (1973). Biological statistics. Minsk: High School. 1973. (In Russian)
6. Snedecor, G.W. (1957). Statistical methods applied to experiments in agriculture and biology. US: The Iowa State college press. 1957.
7. Gubler, E.V., Genkin, A.A. (1973). Application of nonparametric statistical criteria in biomedical research. Leningrad: Medicine. 1973. (In Russian)
8. Prokhorov, Yu.V. (1999). Probability and mathematical statistics: Encyclopedia. Moscow: The Great Russian Encyclopedia. 1999. (In Russian)
9. Hollander, M., Wolfe, D.A. (1999). Nonparametric statistical methods. New York: Wiley-Interscience. 1999.
10. Hettmansperger, T.P. (1991). Statistical inference based on ranks. New York: Krieger Pub Co. 1991.
11. Kendall, M.G. (1962). Rank correlation methods. Chicago: Hafner Publishing Company. 1962.
12. Gibbons, J., Chakraborti, S. (2003). Nonparametric statistical inference. Fourth Edition, Revised and Expanded. New York: Marcel Dekker. 2003.
13. Lehmann, E.L. (1975). Nonparametric statistical methods based on ranks. San Francisco, Düsseldorf etc.: McGraw-Hill. 1975.
14. Puri, M.L., Sen, P.K. (1985). Nonparametric methods in general linear models. New York etc.: Wiley. 1985.
15. Pratt, J.W., Gibbons, J.D. (1981). Concepts of nonparametric theory. New York: Springer-Verlag. 1981. <https://doi.org/10.1007/978-1-4612-5931-2>
16. Sidak, Z., Sen, P.K., Hajek, J. (1999). Theory of rank tests. US: Academic Press. 1999.
17. Tyurin, Yu.N., Makarov, A.A. (1998). Statistical analysis of data on a computer. Moscow: Infra-M. 1998. (In Russian)
18. Tyurin, Yu.N., Shmerling, D.S. (2004). Nonparametric methods of statistics. *Sociology: Methodology, Methods, Mathematical models*, 18, 154–166. (In Russian)
19. Siegel, S. (1957). Nonparametric statistics. *The American Statistician*, 11 (3), 13–19. <https://doi.org/10.1080/01621459.1957.10501392>
20. 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. <https://doi.org/10.1214/aoms/1177730491>
21. Rosenbaum, S. (1954). Tables for a nonparametric test of location. *Annals of Mathematical Statistics*, 25 (1), 146–150. <https://doi.org/10.1214/aoms/1177728854>
22. Wilcoxon, F. (1945). Individual Comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83. <http://www.jstor.org/stable/3001968>.
23. Wilcoxon, F. (1947). Probability tables for individual comparisons by ranking methods. *Biometrics*, 3(3), 119–122. <https://doi.org/10.2307/3001946>
24. Wilcoxon, F. (1946). Individual comparisons of grouped data by ranking methods. *Journal of Economic Entomology*, 39, 269. <https://doi.org/10.1093/jee/39.2.269>
25. Stepanov, V.G. (2019). Application of nonparametric statistical methods in agricultural biology and veterinary medicine research. St-Petersburg: Lan. 2019. (In Russian)
26. Edelbaeva, N.A., Lebedinskaya, O.G., Kovanova, E.S., Tenetova, E.P., Timofeev, A.G. (2019). Fundamentals of nonparametric statistics. Moscow: YUNITY-DANA. 2019. (In Russian)
27. Kuz'mina, E.V., P'yankova, N.G., Tret'yakova, N.V., Botsoeva, A.V. (2020). Using data analysis methodology to foster professional competencies in business informaticians. *European Journal of Contemporary Education*, 9(1), 54–66. <https://doi.org/10.13187/ejced.2020.1.54>
28. Podrihalo, O., Savina, S., Podrigalo, L., Iermakov, S., Jagiello, W., Rydzik, Ł. et al. (2020). Influence of health-related fitness on the morphofunctional condition of second mature aged women. *International Journal of Environmental Research and Public Health*, 17(22), Article 8465, 1–9. <https://doi.org/10.3390/ijerph17228465>
29. Kotlyarova, I., Chuvashova, A. (2020). Educational imitation of basic job function using the knowledge of English among technical Major students. *International Journal of Instruction*, 14(1), 303–324. <https://doi.org/10.29333/IJI.2021.14118A>
30. Mueller, S. K., Nocera, A. L., Dillon, S. T., Libermann, T. A., Wendler, O., Bleier, B. S. (2019). Tissue and exosomal serine protease inhibitors are significantly overexpressed in chronic rhinosinusitis with nasal polyps. *American Journal of Rhinology and Allergy*, 33(4), 359–368. <https://doi.org/10.1177/1945892419831108>
31. Gilic, F., Schultz, K., Sempowski, I., Blagojevic, A. (2019). "Nightmares-family medicine" course is an effective acute care teaching tool for family medicine residents. *Simulation in Healthcare*, 14(3), 157–162. <https://doi.org/10.1097/SIH.0000000000000355>
32. Zabetian, H., Sadeghi, F., Falah, A., Kalani, N. (2018). Studying the Effect of Intravenous Injections of Ketorolac (IVIK) on analgesia control before and after using tourniquet in orthopedic surgery of femur and tibia by general anesthesia. *Journal of Research in Medical and Dental Science*, 6(2), 227–232. <https://doi.org/10.24896/jrmds.20186235>
33. Spielberger, J., Heid, F., Schmidtmann, I., Drees, P., Betz, U., Schwaderlapp, W. et al. (2021). Patient-centered perioperative vigilance: perioperative process quality, effectiveness of pain treatment and mobilization progress after implementation of a treatment bundle for total knee endoprosthesis. *Anaesthesia*, 70(3), 213–222. <https://doi.org/10.1007/s00101-020-00874-8>
34. Regnery, S., Behl, N. G. R., Platt, T., Weinfurter, N., Windisch, P., Deike-Hofmann, K. et al. (2020). Ultra-high-field sodium MRI as biomarker for tumor extent, grade and IDH mutation status in glioma patients. *NeuroImage: Clinical*, 28, Article 102427. <https://doi.org/10.1016/j.nicl.2020.102427>
35. McDonald, J.H. (2014). G-test of goodness-of-fit. Baltimore, Maryland: Sparky House Publishing. 2014.
36. McDonald, J.H. (2014). Small numbers in chi-square and G-tests. Baltimore, Maryland: Sparky House Publishing. 2014.
37. Mangiaterra, G., Amiri, M., Di Cesare, A., Pasquaroli, S., Manso, E., Cirilli, N., Citterio, B. et al. (2018). Detection of viable but non-culturable *Pseudomonas aeruginosa* in cystic fibrosis by qPCR: a validation study. *BMC Infectious Diseases*, 18, Article 701. <https://doi.org/10.1186/s12879-018-3612-9>
38. Bilello, L.A., Scuderi, C., Haddad, I.C.J., Smotherman, C., Shahady, E. (2018). Practice transformation: using team-based care training to improve diabetes outcomes. *Journal of primary care and community health*, 9. <https://doi.org/10.1177/2150132718817952>
39. Taetzsch, A., Das, S. K., Brown, C., Krauss, A., Silver, R. E., Roberts, S. B. (2018). Are gluten-free diets more nutritious? An

- evaluation of self-selected and recommended gluten-free and gluten-containing dietary patterns. *Nutrients*, 10(12), Article 1881. <https://doi.org/10.3390/nu10121881>
40. Raurich, J. M., Llompert-Pou, J. A., García-de-Lorenzo, A., Buño Soto, A., Marsé, P., Frontera, G. et al. (2018). Effect of the route of nutrition and l-alanyl-l-glutamine supplementation in amino acids' concentration in trauma patients. *European Journal of Trauma and Emergency Surgery*, 44(6), 869–876. <https://doi.org/10.1007/s00068-017-0851-1>
41. Runge, T.M., Jirapinyo, P., Chan, W.W., Thompson, C.C. (2019). Dysphagia predicts greater weight regain after Roux-en-Y gastric bypass: a longitudinal case-matched study. *Surgery for obesity and related diseases*, 15(12), 2045–2051. <https://doi.org/10.1016/j.soard.2019.06.041>
42. Lee, D., May, K., Faramarzi, B. (2019). Comparison of first and second acupuncture treatments in horses with chronic laminitis. *Iranian Journal of Veterinary Research*, 20(1), 9–12.
43. Avena, V., Messina, D., Corte, C., Mussi, J., Saez, A., Boarelli, P. et al. (2019). Association between consumption of yerba mate and lipid profile in overweight women. *Nutricion hospitalaria*, 36(6), 1300–1306. <https://doi.org/10.20960/nh.02599>
44. Reaver, A., Hewlings, S., Westerman, K., Blander, G., Schmeller, T., Heer, M. et al. (2019). A randomized, placebo-controlled, double-blind crossover study to assess a unique phytosterol ester formulation in lowering LDL cholesterol utilizing a novel virtual tracking tool. *Nutrients*, 11(9), Article 2108. <https://doi.org/10.3390/nu11092108>
45. Shi, L., Rong, Y., Daly, M., Dyer, B., Benedict, S., Qiu, J. et al. (2020). Cone-beam computed tomography-based delta-radiomics for early response assessment in radiotherapy for locally advanced lung cancer. *Physics in Medicine and Biology*, 65(1), Article 015009. <https://doi.org/10.1088/1361-6560/ab3247>
46. Shauly-Aharonov, M. (2020). An exact test with high power and robustness to unmeasured confounding effects. *Statistics in Medicine*, 39(8), 1041–1053. <https://doi.org/10.1002/sim.8460>
47. Mou, T., Deng, W.J., Gu, F.Y., Pawitan, Y., Vu, T.N. (2020). Reproducibility of methods to detect differentially expressed genes from single-cell RNA sequencing. *Frontiers in Genetics*, 10, Article 1331. <https://doi.org/10.3389/fgene.2019.01331>
48. Sasaki, C.A.L., da Costa, T.H.M. (2021). Micronutrient deficiency in the diets of para-athletes participating in a sports scholarship program. *Nutrition*, 81, Article 110992. <https://doi.org/10.1016/j.nut.2020.110992>
49. Anderson, D., Sturt, J., McDonald, N., White, C., Porter-Steele, J., Rogers, R. et al. (2021). International feasibility study for the women's wellness with type 2 diabetes programme (WWDP): An eHealth enabled 12-week intervention programme for midlife women with type 2 diabetes. *Diabetes Research and Clinical Practice*, 171, Article 108541. <https://doi.org/10.1016/j.diabres.2020.108541>

## AUTHOR INFORMATION

**Marina A. Nikitina** — candidate of technical sciences, docent, leading scientific worker, the Head of the Direction of Information Technologies of the Center of Economic and Analytical Research and Information Technologies, V. M. Gorbатов Federal Research Center for Food Systems of Russian Academy of Sciences. 26, Talalikhina, 109316, Moscow, Russia. Tel: +7-495-676-95-11 extension 297, E-mail: m.nikitina@fncps.ru  
ORCID: <https://orcid.org/0000-0002-8313-4105>

\* corresponding author

**Irina M. Chernukha** — doctor of technical sciences, professor, Academician of the Russian Academy of Sciences, Head of the Department for Coordination of Initiative and International Projects, V. M. Gorbатов Federal Research Center for Food Systems of Russian Academy of Sciences. 26, Talalikhina, 109316, Moscow, Russia. Tel: +7-495-676-95-11 extension 109, E-mail: imcher@inbox.ru  
ORCID: <https://orcid.org/0000-0003-4298-0927>

All authors bear responsibility for the work and presented data.

All authors made an equal contribution to the work.

The authors were equally involved in writing the manuscript and bear the equal responsibility for plagiarism.

The authors declare no conflict of interest.