International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 5, No. 3, 2024

Methods for Testing the Significance of Differences in Biological Testing

Baoqin Jiang *, Dejie Feng, Jing Ren, Chanchan An, Yan Liu, Qinqin Wei, Huanhuan Zheng,

Xin Chen, Hongyan Jin

Lanzhou Institute of Biological Products Co. Ltd., Lanzhou Gansu, 730046, China

*** Corresponding author:** Baoqin Jiang

Abstract: Bioassays play a crucial role in the research of new drugs and vaccines. The significance test for differences is the most commonly used statistical method in the analysis of biometric data. The methods for testing the significance of differences in paired samples can be broadly divided into two categories. The first category is that the data population follows a normal distribution, and the commonly used method is paired t-test; The second type requires the use of nonparametric testing methods for statistical testing when parameter testing methods are not applicable. Commonly used methods include Wilcoxon matched signature level testing. This article introduces the basic principles and prerequisites of two types of testing methods, and in establishing a detection method for trypsin residue, takes the significance of the difference between the results of three enzyme-linked immunosorbent assay (ELISA) tests at 36℃ and the results at 37℃ for two batches of raw materials as an example to explain the correct use of the two testing methods. Correctly applying significance testing can establish the conclusions of experiments or investigations on a more scientific and reliable basis, avoiding simplification and absolutization.

Keywords: Significance of Differences; Paired T-test; Wilcoxon Matching Signature Level Test; Normal Distribution; Bioassay.

1. Introduction

Biological testing is an important content and foundation in biology, medicine, and other fields. In the process of researching new drugs and vaccines, biological testing plays a crucial role. Significance testing, also known as hypothesis testing, is a statistical inference method used to determine whether the differences between samples or between samples and populations are caused by sampling errors or essential differences. The basic principle is to first make a certain assumption about the characteristics of the population, and then infer whether to accept this assumption through statistical analysis of sampling research. If accepted, i.e. the difference between populations is not statistically significant, it is determined that the difference is caused by sampling or measurement errors; If rejected, that is, if the differences between populations are statistically significant, it is determined that the differences are caused by research factors. The significance test for differences is the most commonly used statistical method in the analysis of biometric data. Usually, we need to compare whether there is a significant difference between two groups of data, and also identify the variable of difference between different groups based on the significance test. This requires the use of statistical hypothesis testing methods to test for differences between groups and calculate their degree of difference[1-3]

In mathematical statistics, a probability (P) of 5% is generally used as the significant evaluation criterion, which means that in 100 trials, if the likelihood of differences caused by accidental factors is more than 5 times, the difference is considered insignificant. If the difference between the two is within a probability range of 5%, and the chance of such a probability occurring is very small, then we consider this difference to have a significant degree of difference. Sometimes we think that 5% is too low and can be raised to 1% as a significant evaluation criterion. If the difference between the two is within a probability range of 1%, then we consider the difference to be extremely significant [4, 5]

Two sets of samples that are interrelated belong to paired samples, such as two sets of values for a certain indicator before and after intervention, which belong to self paired samples. There are generally two types of methods for testing the significance of differences in paired samples. The first type is parameter testing, which is based on the observation data of the sample to test the differences in population parameters (such as population mean and variance) and population parameters. If the data population follows a normal distribution, the commonly used method is paired t-test. T-test based on t-distribution theory is convenient to calculate and has high testing power, and is the most used biological testing method; The second type is non parametric testing methods. In the actual analysis process, data often encounters situations where the prerequisite conditions for parametric testing are not met, such as the population not following a normal distribution or the population distribution is unknown. In this case, parametric testing methods cannot be used, otherwise the test results obtained will be inaccurate. In cases where parametric testing methods are not applicable, non- parametric testing methods such as Wilcoxon matching signature level testing are commonly used for statistical testing. Therefore, this article introduces the basic principles and prerequisites of the two types of testing methods, and explains their correct usage methods[6-10]

2. Paired t Test

2.1. The Basic Principle of Paired T-test

The t-distribution was first developed by British statistician W S. Gosset published under the pen name "student" in 1908

[11, 12], ushered in a new era of small sample statistical inference. If $X \sim N(0,1)$, $Y \sim \chi_n^2$, and X and Y are independent, then the distribution of a random variable is called a t-distribution with n degrees of freedom (*df*) and is denoted as $t \sim t_n$.

$$t = \frac{X}{\sqrt{Y/n}} \tag{1}$$

The probability density function of the t-distribution is:

$$t_{(x,n)} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \tag{2}$$

In equation (2), Γ (·) is the gamma function.

The test statistic corresponding to paired t-test is:

$$t = \frac{|\bar{d}-0|}{s_{\bar{d}}}, s_{\bar{d}} = \frac{s}{\sqrt{n}}, df = n-1 \tag{3}$$

In equation (3), n is the number of paired children, and d is the difference between paired data.

2.2. Analysis of the Application Conditions of Paired T-test Method

Data satisfaction is the basis for conducting t-tests. The t-test is a statistical method for comparing the differences between two groups of data that conform to the t-distribution. Further attention should be paid to the requirements of data characteristics, that is, when comparing means and considering from the perspective of the t-distribution, there are two prerequisites for paired t-tests, namely independence and normality.

2.2.1. Test of Independence

Independence refers to the mutual independence between various observation values, which can be judged based on professional knowledge or common sense. The paired design of quantitative data requires that the measurement values between different pairs should be independent of each other. Taking the establishment of a detection method for trypsin residue as an example, it is necessary to detect the significant difference between the results of three enzyme-linked immunosorbent assay (ELISA) tests at 36℃ and 38℃ and the results at 37℃. The measured values between the samples at 36℃ and 37℃ are independent of each other, and the measured values between the samples at 38℃ and 37℃ are independent of each other, meeting the independence test.

2.2.2. Normal Distribution Test

If one wishes to use t-test to process quantitative data for single group and group designs, the prerequisite is that the quantitative data for each group to be analyzed follows or approximates a normal distribution, or is transformed into a normal distribution through data transformation. For paired design quantitative data, there is no requirement for two sets of raw data, only the difference between the two sets needs to follow a normal distribution. Taking the results of three enzyme-linked immunosorbent assay (ELISA) tests at 36℃ and 37℃ (ng/mL) for a certain batch of raw materials as an example, Table 1 shows the comparative data of ELISA test results under two temperature conditions.

Table 1. Comparison data of enzyme-linked immunosorbent assay results under two temperature conditions

times	36℃	37℃	difference	times	36℃	37℃	difference
1	5.5377	4.8978	0.6399	**21**	5.6715	3.9109	1.7606
2	6.8339	4.7586	2.0753	**22**	3.7925	5.0326	1.2401
3	2.7412	5.3192	2.578	**23**	5.7172	5.5525	0.1647
4	5.8622	5.3129	0.5493	**24**	6.6302	6.1006	0.5296
5	5.3188	4.1351	1.1837	**25**	5.4889	6.5442	1.0553
6	3.6923	4.9699	1.2776	**26**	6.0347	5.0859	0.9488
7	4.5664	4.8351	0.2687	**27**	5.7269	3.5084	2.2185
8	5.3426	5.6277	0.2851	**28**	4.6966	4.2577	0.4389
9	8.5784	6.0933	2.4851	**29**	5.2939	3.9384	1.3555
10	7.7694	6.1093	1.6601	**30**	4.2127	7.3505	3.1378
11	3.6501	4.1363	0.4862	**31**	5.8884	4.3844	1.504
12	8.0349	5.0774	2.9575	**32**	3.8529	5.7481	1.8952
13	5.7254	3.7859	1.9395	**33**	3.9311	4.8076	0.8765
14	4.9369	3.8865	1.0504	**34**	4.1905	5.8886	1.6981
15	5.7147	4.9932	0.7215	**35**	2.0557	4.2352	2.1795
16	4.795	6.5326	1.7376	**36**	6.4384	3.5977	2.8407
17	4.8759	4.2303	0.6456	**37**	5.3252	3.5776	1.7476
18	6.4897	5.3714	1.1183	**38**	4.2451	5.4882	1.2431
19	6.409	4.7744	1.6346	**39**	6.3703	4.8226	1.5477
20	6.4172	6.1174	0.2998	**40**	3.2885	4.8039	1.5154

Table 2 shows the normal distribution detection results of the original data. The input data includes the results and their differences of enzyme-linked immunosorbent assay (ELISA) at 36℃ and 37℃. The experiment was conducted 40 times, and the normality of the input data was tested using four testing methods: Anderson Darling test, D'Agostino&Pearson test, Shapiro Wilk test, and Kolmogorov Smirnov test at the α=0.05 level, all three sets of data passed these four normal

tests and belong to a normal distribution. The probabilities of the three sets of data belonging to Gaussian normal distribution are 91.48%, 82.91%, and 82.32%, respectively. The probabilities of belonging to logarithmic normal distribution are 8.524%, 17.09%, and 17.68%, respectively. The likelihood ratios (LR) are 10.73, 5.4906, and 4.657, respectively. The skewness is 0.02862, 0.3809, and 0.3986, respectively. The kurtosis is 0.2532, -0.2589, and -0.4912, respectively. Based on comprehensive judgment, all three sets of data belong to Gaussian normal distribution. As shown in Figure 1, the Q-Q plot of the difference between the results of enzyme-linked immunosorbent assay (ELISA) at 36℃ and 37℃ in the original solution. From the graph, it can be seen that all points in the three sets of data follow a straight-line y=x, and the sample distribution is similar to the theoretical distribution. There is basically no systematic bias in the data, and the data clearly follows a normal distribution.

Table 2. Normal distribution test results

Comparing normal distribution and lognormal distribution			
Probability Normal (Gaussian)	91.48%	82.91%	82.32%
Probability log normal	8.524%	17.09%	17.68%
Likelihood ratio (LR)	10.73	5.4906	4.657
1/LR	0.09318	0.038	0.2147
Which distribution is more likely?	Normal	Normal	Normal
Testing normal distribution			
Anderson-Darling test			
A2*	0.2794	0.3196	0.313
P-value	0.6281	0.5211	0.5342
Passed normality test (alpha=0.05)?	yes	yes	yes
Summary of P-values	ns	ns	ns
D'Agostino & Pearson test			
K2	0.3335	1.15	1.63
P-value	0.8464	0.5627	0.4427
Passed normality test (alpha=0.05)?	yes	yes	yes
Summary of P-values	ns	ns	ns
Shapiro-Wilk test			
W	0.9856	0.9712	0.9658
P-value	0.8831	0.393	0.2628
Passed normality test (alpha=0.05)?	yes	yes	yes
Summary of P-values	ns	ns	ns
Kolmogorov-Smirnov test			
KS	0.09721	0.08744	0.07784
P-value	>0.1000	>0.1000	>0.1000
Passed normality test (alpha=0.05)?	yes	yes	yes
Summary of P-values	ns	ns	ns

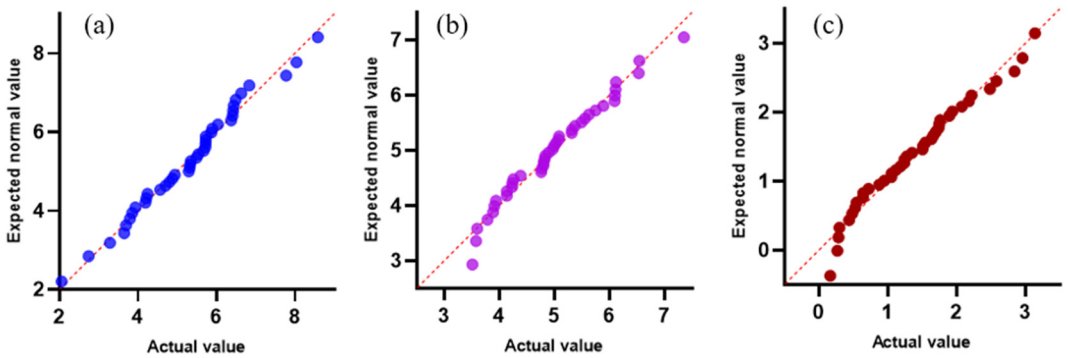

Figure 1. Q-Q plot of the difference between the enzyme-linked immunosorbent assay results at 36℃ and 37℃ in the original solution

2.3. Practical Application of Paired T-test Method

In the establishment of a detection method for trypsin residue, taking the significant difference between the results of three enzyme-linked immunosorbent assay (ELISA) tests at 36℃ and the test results at 37℃ for a certain batch as an example, in the previous section, the data was analyzed to meet independence and normal distribution, following the

"four principles" (random principle, control principle, repetition principle, balance principle) to ensure good representativeness of the sample, and to ensure excellent balance of important non experimental factors among the subjects in each group, thereby improving inter group comparability.

The specific calculation steps for the paired t-test method are as follows:

(1) Calculate the standard error of differences;

(2) Calculate the *t*-value through the average of differences and the standard error of differences;

(3) Obtain the critical *t* value from the degree of freedom *df* value;

(4) Compare the calculated *t*-value with the critical *t*-value and make statistical judgments.

Table 3. Pairing effect analysis

How is the pairing effect?	
Correlation coefficient (r)	0.1053
P-value (single tailed)	0.259
Does pairing have a significant impact?	no

Table 4. Difference results

How big is the difference?	
The average value of differences	-0.3136
Differential SD	1.577
Differential SEM	0.2494
95% confidence interval	[-0.8181 0.1909]
R-squared (partial eta squared)	0.03895

Table 5. Paired t-test results

Paired t-test	
P-value	0.2161
Is there a significant difference (P<0.05)?	no
Single tailed or double tailed P-value?	double tailed
t, *df*	*t*=1.257, *df*=39
Number Pairs	40

The data pairing effect analysis is shown in Table 3. The analysis results indicate that the correlation coefficient between the three enzyme-linked immunosorbent assay (ELISA) results at 36℃ and the results at 37℃ is 0.1053, and the P-value (single tailed) is 0.259. The pairing has no significant effect. Figure 2 shows the heat map of the difference in data between the two, and Table 4 shows the difference results. Through Anderson Darling (A2 *), D'Agostino Pearson omnibus (K2), and Shapiro Wilk (W) tests, the P-values were 0.1609, 0.2307, and 0.2269, respectively, and were tested for normal distribution. The average difference shown in the figure is -0.3136, the SD value is 1.577, the SEM value is 0.2494, and the 95% confidence interval is -0.8181 to 0.1909.

Table 4 shows the paired t-test results, and Figure 3 shows the paired t-test results graph. The paired t-test showed a two tailed P-value of 0.2161, with P>0.05 as the standard. There was no significant difference between the results of three enzyme-linked immunosorbent assay (ELISA) tests at 36 ℃ and those at 37 ℃.

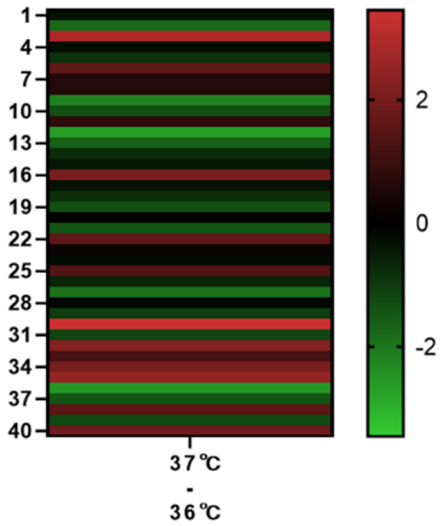

Figure 2. Differential heat map

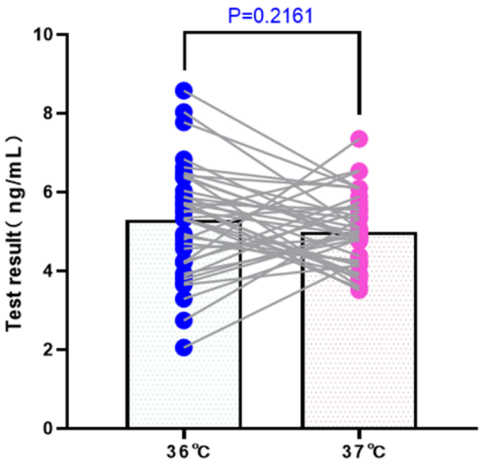

Figure 3. Paired t-test results

3. Wilcoxon Matched Paired Signature Level Test

If the data belongs to a special distribution or has certain characteristics, it can be transformed into normality or homogeneity of variance after certain transformations, and then tested using paired t-test. If the data conforms to the Poisson distribution, it can be transformed by square root; The data that conforms to the binomial distribution can be transformed using the square root inverse sine function; It can also be converted through logarithmic transformation. However, when the overall distribution of two sets of data cannot be determined or there is no appropriate conversion method, non parametric statistical methods can be used. Non parametric statistical methods compare distributions rather than parameters. It does not consider the distribution type of the data, but directly compares it using the symbol, size order number, comprehensive judgment of the ranking, severity, or quality level of the sample data. The level data that is difficult to process by the parametric method can be analyzed by the non parametric method, so its application range is wide[2, 6, 9]. This article focuses on analyzing the Wilcoxon matching signature level testing method.

3.1. Basic Principles

The Wilcoxon paired signature level test is a commonly used non parametric test method used to compare the differences between two related samples. The basic principle

of Wilcoxon matching signature level test is to take the absolute value of the difference between two related samples, arrange them according to the size of the absolute value, assign a rank to each difference, and finally use the sum of ranks as the test statistic. If the difference distribution between two samples is consistent, then the sum of their ranks should be similar; If the difference distribution between two samples is not consistent, the sum of their ranks will have a significant difference.

The hypothesis of Wilcoxon's paired signature level test is as follows:

Zero hypothesis (H0): The difference distribution between two samples is consistent.

Alternative hypothesis (H1): The difference distribution between two samples is inconsistent.

According to the hypothesis test results, if the p-value is less than the significance level (usually 0.05), the null hypothesis is rejected, and it is considered that the difference distribution between the two samples is inconsistent.

3.2. Practical Application of Wilcoxon Matched Paired Signature Level Test

In the establishment of a detection method for trypsin residue, taking the significant difference between the results of three enzyme-linked immunosorbent assay (ELISA) tests at 36℃ and the results at 37℃ for a certain batch of raw liquid as an example. The data is shown in Table 6.

Table 6. Three enzyme-linked immunosorbent assay (ELISA) results of a batch at 36℃ and 37℃

times	1	2	3	4	5	6	7	8
37℃	5.44244	5.54558	6.11553	6.43886	5.91751	5.04235	5.32146	5.39185
36℃	5.40983	5.22338	5.65287	6.70279	6.39517	6.70681	6.71691	9.38234
difference	0.03261	0.3222	0.46266	0.26393	0.47766	1.66446	1.39545	3.99049

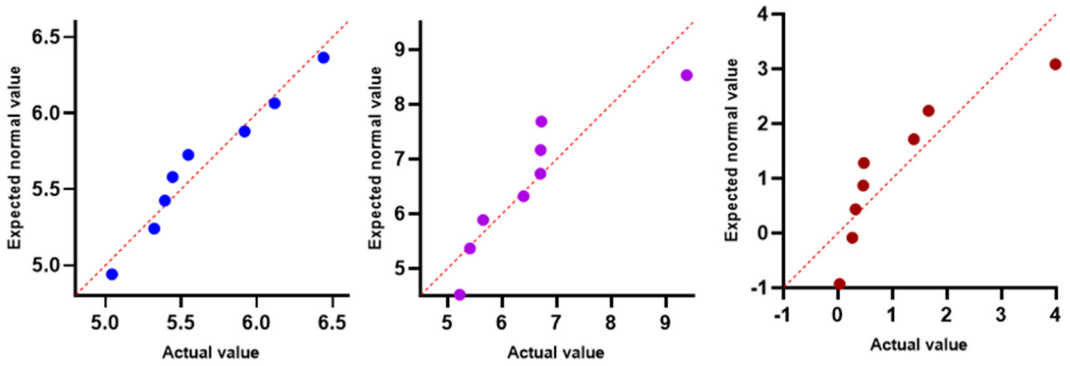

Figure 4. Q-Q plot of the difference between the enzyme-linked immunosorbent assay results at 37℃ and 36℃ in the original solution

Figure 4 shows the Q-Q plot of the enzyme-linked immunosorbent assay results and their differences under the conditions of 37℃ and 36℃ as input data. The experiment was conducted 8 times, and the normality of the input data was tested using four testing methods: Anderson Darling test, D'Agostino& Pearson test, Shapiro Wilk test, and Kolmogorov Smirnov test at the α=0.05 level, the 37℃ data all passed these four normal tests and belong to a normal distribution. However, the 36℃ and difference data did not pass the normal test, and the probabilities of the three sets of data belonging to Gaussian normal distribution were 31%, 46.62%, and 1.321%, respectively. Based on comprehensive judgment, the 36℃ data basically belongs to Gaussian normal distribution, but the 36℃ and difference data do not belong to Gaussian normal distribution. This article uses Wilcoxon matching signature level test for non parametric testing.

Table 7 shows the median difference results, and Figure 5 shows the difference between the two data. The median difference is 0.3708, and the 99.22% confidence interval is [-0.4627 3.990]. Table 8 shows the results of the Wilcoxon paired signature level test, and Figure 6 shows the results of the Wilcoxon paired signature level test. The results showed that its double tailed P value was 0.1953, and with P>0.05 as the standard, there was no significant difference in the results of three enzyme-linked immunosorbent assay (ELISA) tests under 36℃ and 37℃ conditions.

Table 7. Difference median

Difference median	
Median	0.3708
99.22% confidence interval	[-0.4627 3.990]

Table 8. Wilcoxon matched paired signature level test results

Wilcoxon Matched Paired Signature Level Test	
P-value	0.1953
What is the exact or approximate P-value?	accurate
Is there a significant difference (P<0.05)?	no
Single tailed or double tailed P-value?	double tailed
The sum of positive and negative ranks	28.00, -8.000
Total signed rank (W)	20
Logarithm	8

4. Discussion

The application of t-test has clear limitations, and blind use of t-test will reduce the reliability of conclusions and even lead to incorrect conclusions. Before conducting the test, it is necessary to analyze the application conditions of the data, with a focus on analyzing the independence and normal distribution of the data. On this basis, it can be decided whether to use parametric or nonparametric tests. Correctly applying significance testing can establish the conclusions of

experiments or investigations on a more scientific and reliable basis, avoiding simplification and absolutization.

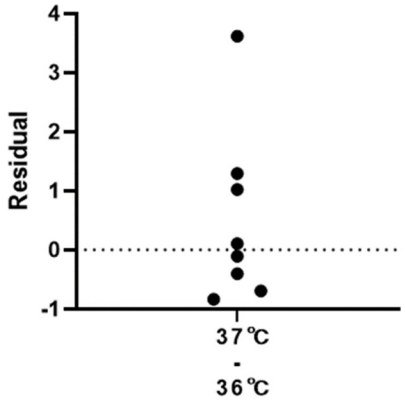

Figure 5. Difference Diagram

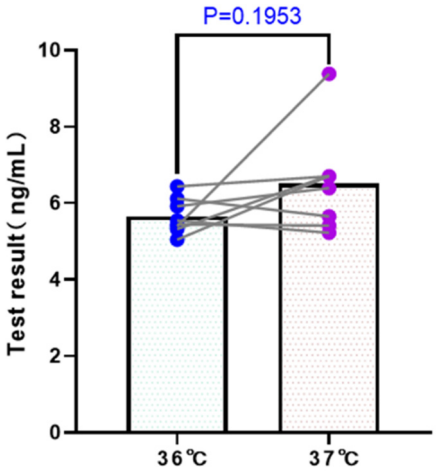

Figure 6. Test Results

5. Conclusion

The significance test for differences is the most commonly used statistical method in the analysis of biometric data, but it requires a focus on its usage conditions.

(1) Data satisfaction is the basis for conducting t-tests. Focus on the prerequisites for testing, namely independence and normality.

(2) Under the premise of ensuring the independence and normality of the analysis data, use paired t-test to test the significance of the difference between the two groups of data;

(3) If the independence and normality of the data are not satisfied, and cannot be converted to a normal distribution through transformation, Wilcoxon matching signature level test is used to test the significance of the difference between the two sets of data.

References

[1] West RM. Best practice in statistics: Use the Welcht-test when testing the difference between two groups. Annals of Clinical Biochemistry. 2021;58(4):267-9.

[2] Mishra P, Pandey C, Singh U, Gupta A, Keshri A. Descriptive statistics and normality tests for statistical data. Annals of Cardiac Anaesthesia. 2019;22(1):67-72.

[3] Vieira PCDC. T-Test with Likert Scale Variables. Social Science Electronic Publishing. 2016.

[4] Trajkovski VE. HOW TO SELECT APPROPRIATE STATISTICAL TEST IN SCIENTIFIC ARTICLES. Faculty of Philosophy, Institute of Special Education. 2016;(17).

[5] Spottiswoode CN, Olsson U, Mills MSL, Cohen C, Francis JE, Toye N, et al. Rediscovery of a long-lost lark reveals the conspecificity of endangeredHeteromirafrapopulations in the Horn of Africa. Journal of Ornithology. 2013;154(3):813-25.

[6] Takada Y, Shiotome N. ROBUSTNESS OF A TWO-STAGE ESTIMATION PROCEDURE WHEN VARIANCES ARE UNEQUAL. Sequential Analysis. 2012;34(3):336-49.

[7] Ruxton GD. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. Behavioral Ecology. 2010;17(4):688-90.

[8] Neuh, aauml, ser, Markus. Two-sample test when variances are unequal. Animal Behaviour. 2002.

[9] Ugoni A, Walker B. THE t TEST: An Introduction. 1995;4.

[10] Murphy BP. Some two-sample tests when the variances are unequal: a simulation study. Biometrika. 1967;(3-4):3-4.

[11] Green KCE. Analysis of Variance: Is There a Difference in Means and What Does It Mean? Journal of Surgical Research. 2008.

[12] Stroes-Gascoyne S, Schippers A, Schwyn B, Poulain S, Sergeant C, Simonoff M, et al. Microbial Community Analysis of Opalinus Clay Drill Core Samples from the Mont Terri Underground Research Laboratory, Switzerland. Geomicrobiology Journal. 2007;24(1):1-17.