UDC 004.93

LabVIEW STATISTICAL DATA ANALYSIS SYSTEM FOR EVIDENCE-BASED MEDICINE

Solomin A.V., PhD, as.prof.
andr-sol@i.ua
Ahmed Sani, student
ahmed.shahahmedshah@hotmail.com
Faculty of Biomedical Engineering
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine

Abstract *– The problem of creating tools to simplify the correct use of statistical research methods remains relevant, especially in the biomedical field, where experts do not have sufficient mathematical background. The article proposes the implementation of a semi-automated system in the NI LabVIEW software environment, which implements the most commonly used statistical analysis tools, but provides verification of the necessary criteria and limits of their use, as well as selection of appropriate statistical methods in automatic mode. Implementation in the software environment NI LabVIEW provides easy connection to the equipment, versatility and ease of further modification of algorithms.*
Keywords: *statistical methods, evidence-based medicine, reliability of conclusions, error probability, application criteria, tree structure, analysis of variance, NI LabVIEW*

I. INTRODUCTION

One of the main methods of obtaining scientifically reliable data in the biomedical sector are evidence-based medicine. It is based on statistical methods of information analysis. But, according to some studies [1-4], a lot of conclusions made by this method are erroneous. This is due to the complexity of statistical analysis mathematical apparatus, primarily the criteria and limits of possible use of statistical tools, and the insufficient level of their mastery of biomedical specialists.

There are a number of computer software packages that facilitate the use of statistical methods of information processing, but the problem remains, because for correct use of specific tools from these packages, professionals still need to know quite deeply and check the necessary conditions.

It should be emphasized that the consequences of erroneous conclusions in evidence-based medicine may be quite dangerous because they relate to human health. Reducing the likelihood of errors is very important.

II. THE AIM OF THE STUDY

The aim of the work was to create a software system that implements the most commonly used tools of statistical analysis, but at the same time provides verification of necessary criteria and limits of their use in automatic mode. Thus, user is not required to know appropriate mathematical apparatus, the system will automatically select optimal statistical methods and formulate conclusions according to specific conditions, significantly reducing the chance of errors.

III. MATERIALS AND METHODS

NI LabVIEW was chosen as the software development environment, which is currently an informal standard in the field of biomedical instrumentation and biomedical research and is easily integrated into most modern software and hardware systems [5-7].

NI LabVIEW (National Instrument Laboratory Virtual Instrument Engineering Workbench) is very convenient for applications that are created and used not by programmers but by engineering professionals, including biomedical engineers. Instead of text programming languages with complex syntactic rules, the graphic language G is used here, which has a more familiar to engineers type of block diagrams.

Additional advantages of the workbench are the presence of a large number of built-in functions and routines, a wealth of developed and visual elements for building interfaces, the ability to interact with other environments.

The convenience of combining with hardware for entering information provides opportunities to create automated workstations for statistical processing, and ease of programming – the ability to improve, configure and adapt the software system to specific situations.

To achieve this goal, a computer software system was built in the form of a tree, where the nodes automatically check the conditions and criteria for specific statistical methods, branching to the next nodes of the algorithm, and the final node of chain draws statistical conclusions.

The result of the software system is formation of a conclusion about existence or absence of a statistically significant difference between average values of two data sets and the statistical relationship between them. In practice, this may relate, for example, to the detection of statistically significant effects of a pharmacological agent, drug procedure, regimen and other factors [8-10].

The sequence of execution of system modules is as follows:

- in module 1 branching occurs into module 2 or module 7, depending on whether the samples presented in the arrays are independent or interconnected; it is implied that either influence of the factor is studied on the experimental and control group, or on one group – before and after influence of the factor;
- in module 2 the normality of the law of data distribution in arrays is checked; if both samples are distributed according to the normal law, then transition occurs to module 3, otherwise – to module 6;
- the module 3 checks uniformity of variances of the input samples by Fisher's test; depending on the result, it is further proposed to branch to module 4 (testing the hypothesis of equality of mean values of two samples by Student's t-test for equal variances) or to module 5 (testing the hypothesis of equality of average values of two samples by Student's t-test for different variances);
- in module 6 the test of hypothesis about the equality of average values of two samples with a distribution law other than normal is implemented by ranking the data and applying the two-sample Wilcoxon test;
- in module 7 the form of distribution law of both input arrays is checked (for the case of connected samples) – is the distribution normal or not. Depending on the result, it is proposed to move to module 8 (testing the hypothesis of equality of mean values of two related samples by Student's t-test) or to module 9 (application of non-parametric sign criterion);
- in module 10, the existence of a parametric correlation between input arrays is checked, if the distribution law is normal, and the Pearson's correlation coefficient is calculated;
- in module 11, the existence of a rank correlation between input arrays is checked if the distribution law is different from normal, and the Spearman's rank correlation coefficient is calculated;
- in module 12 the one-factor analysis of variance of input arrays is implemented in order to detect the influence of the factor, that is to identify a statistically significant difference between average values of the samples by means of analysis of variance.

It should be noted that at each node of the tree structure, the conditions for applicability of statistical methods are checked before transition to the corresponding module occurs. And in each module, the statistical accuracy of the conclusion is checked.

For example, in correlation analysis, not only the correlation coefficient is calculated, but also the statistical accuracy of such a correlation is determined. The appearance of the system interface is shown in Fig.1.

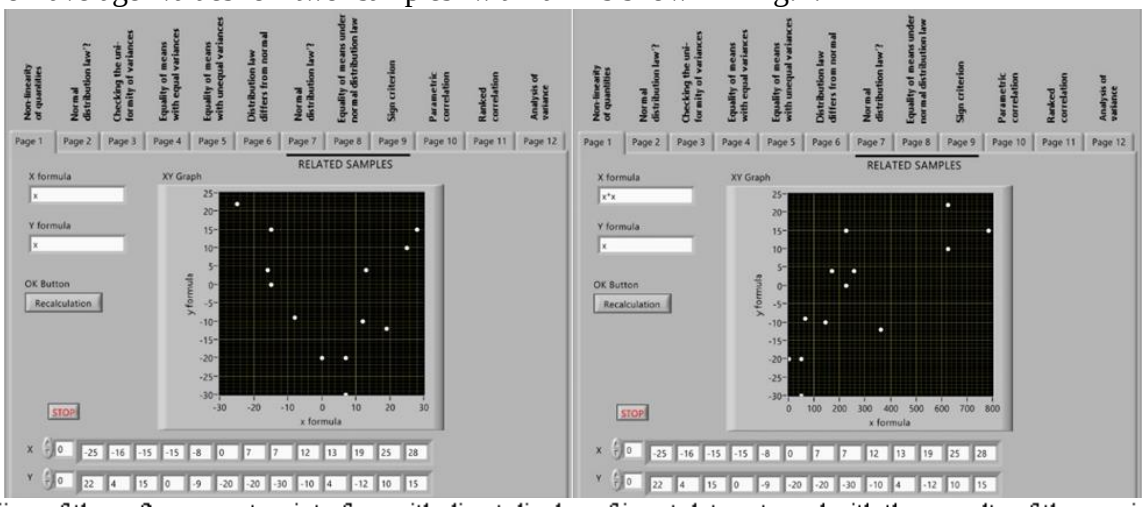

Fig. 1. View of the software system interface with direct display of input data sets and with the results of the previous functional transformation

Here at the top there are 12 pages of the interface (bookmarks), which correspond to execution of the above modules and are sequentially opened during execution of the algorithm. At the bottom – the input data arrays, given for illustration and settings, and in real circumstances can be entered automatically from various devices or software.

Two options for displaying input data on the left and right screens demonstrate a specially designed ability of the system to analyze either the input data arrays directly, or the results of their preliminary conversion by almost any function that can be simply entered in the interface window. In the example shown in Fig. 1 on the first screen you can see signs that between the input data arrays there is a relationship similar to quadratic law, so on the second screen instead of the data of one of the arrays enter their squares, then see the trend of linearity of such data connection law. It is possible to select functions of preliminary transformation of input data either on a kind of curves on the screen, or proceeding from physical essence of the investigated phenomena.

The result of each module is either a branch demonstration with an explanation of reasons for a particular transition, or a statistical conclusion of an algorithm. Thus, after selecting the functions of homomorphic transformation in the first module, a branching take place due to the choice of the option regarding the independence of samples or their connection. Next, the normality of the distribution law of both samples is checked, then the equality of variances; each combination of features uses its own statistical tool. Checks in the modules are performed automatically using the appropriate statistical criteria.

An example of display in the interface of one of the statistical conclusions (analysis of variance) is shown in Fig.2.

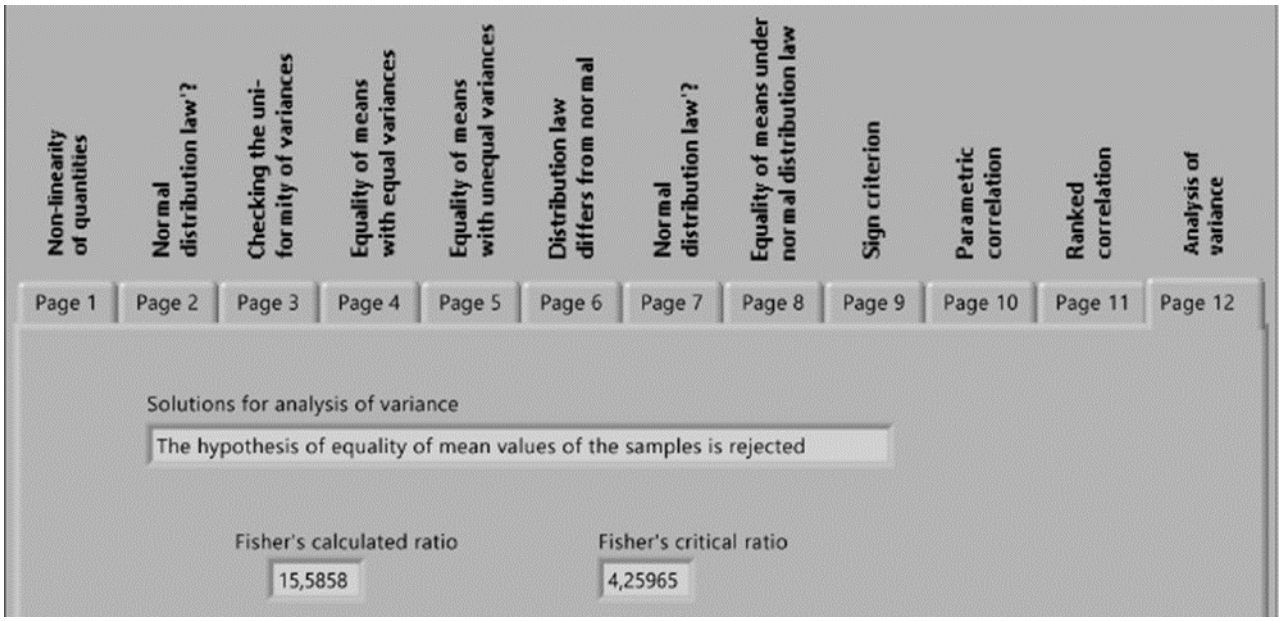

Fig. 2. Display in the interface of t he statistical conclusion of analysis of variance

Here in the text line it is stated that “The hypothesis of equality of mean values of the samples is rejected”, and in the numerical indicators the values of Fisher's calculated ratio and Fisher's critical ratio are given, on which basis this statistical conclusion is made.

For example, let’s consider the structure and operation principle of several modules of the constructed system.

IV. RESULTS AND DISCUSSION

Block diagram module that implements the analysis of variance, shown in Figure 3

The module calculates actual Fisher ratio MSa/ MSe – the ratio of variance, which is explained by the influence of factor (intergroup), and unexplained variance (intragroup), and compares with the critical value of Fisher for the appropriate number of degrees of freedom and a given level of significance. The calculated indicators are displayed on the front panel, and based on the result of the comparison, a statistical conclusion is made about the acceptance or rejection of the hypothesis about equality of the arrays average values. This output is also displayed on the interface panel.

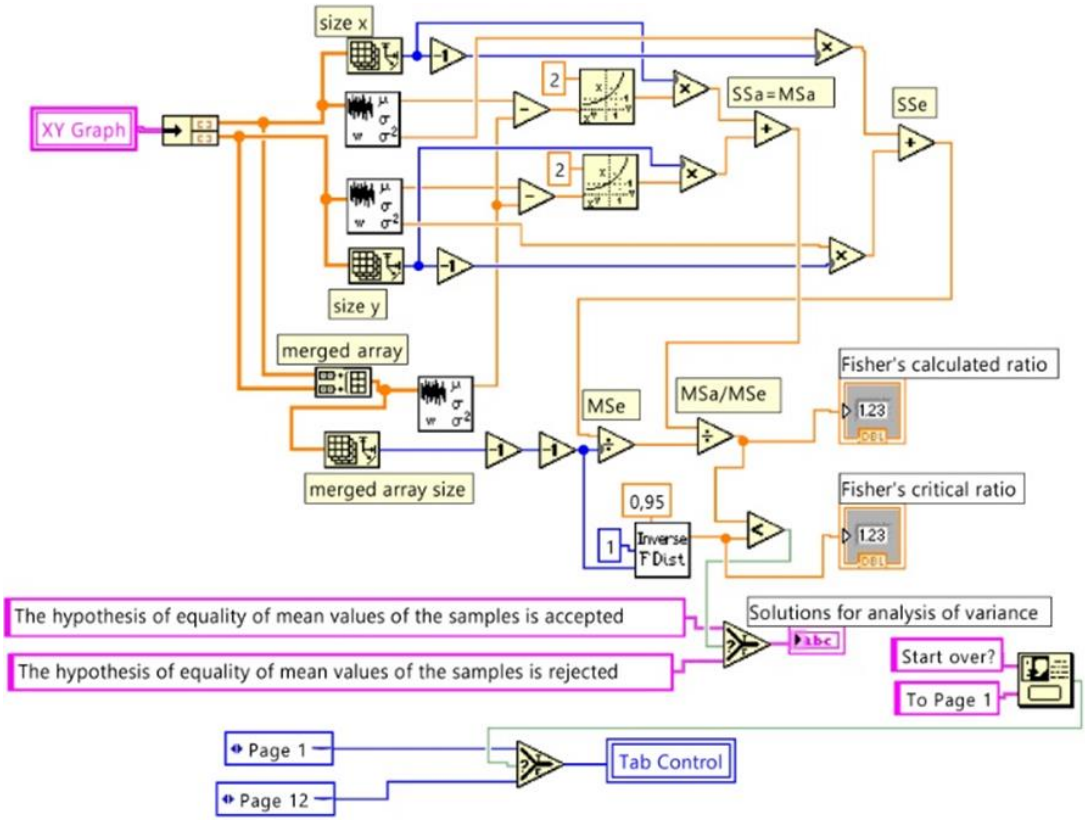

Fig. 3. Block diagram of the module for analysis of variance

It should be noted that the module of analysis of variance seems to duplicate the results of previous modules, i.e. determines the presence of a statistically significant difference between the average values of two samples, only by another method. But, first, the verification of important conclusions by two independent statistical methods increases the reliability of obtained results, and secondly, the analysis of variance is less sensitive to errors associated with a small amount of data in the samples.

Block diagrams of modules that perform correlation analysis for two situations – the normal law of sample distribution and the law other than normal, are shown, respectively, in Fig.4 and Fig.5.

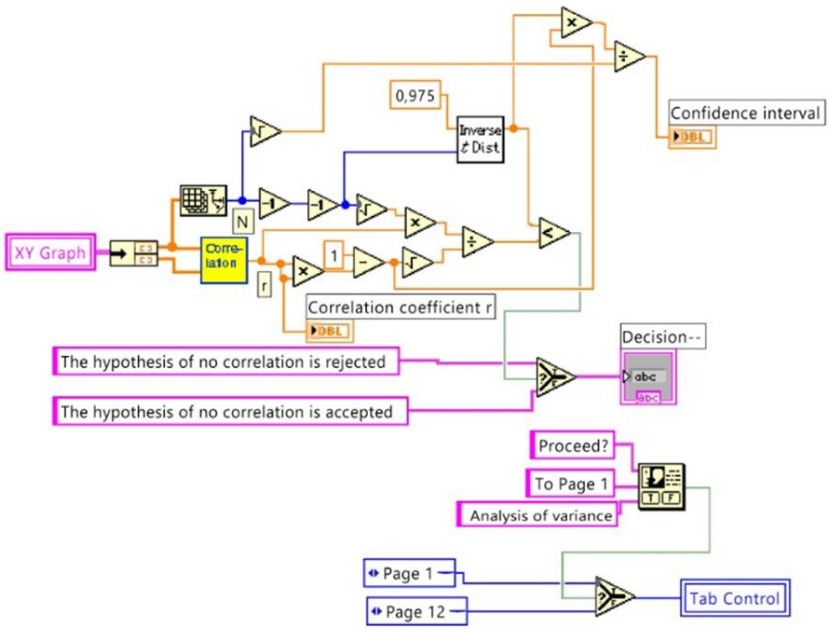

Fig. 4. Block diagram of the correlation analysis module for samples distributed according to the normal law

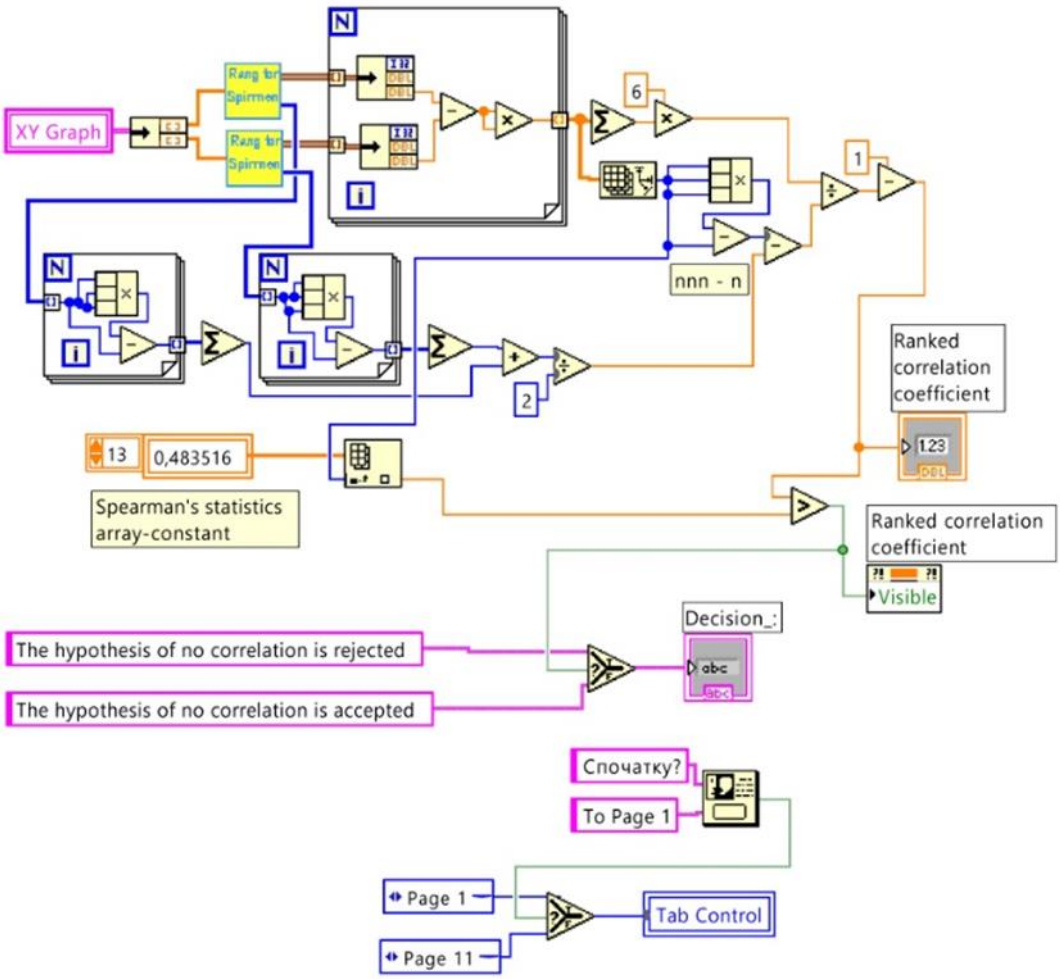

Fig. 5. Block diagram of the correlation analysis module for samples distributed according to law other than normal

In the module of Fig. 4 the detection of parametric correlation is realized and the Pearson correlation coefficient is calculated by the formula

$$r_{xy} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \sum_{i=1}^{N}(Y_i - \overline{Y})^2}}$$

But then the statistical accuracy of the result is checked, i.e. the hypothesis of equality of the correlation coefficient to zero is tested. The interface displays conclusion about the presence or absence of a correlation, the value of correlation coefficient and the half-width of confidence interval for it.

In the module of Fig. 5 the detection of rank correlation is realized and the Spearman rank correlation coefficient is calculated by the formula

$$\rho(A,B) = 1 - \frac{6\sum_{i=1}^{n}(R_{1i} - R_{2i})^2}{n^3 - n},$$

where R_{1i} and R_{2j} are the ranks of respective objects of two samples.

After that, the significance of result is checked according to special tables of Spearman's statistics. This module also uses the developed subroutine Rangir for Spirmen, which performs the ranking of input data sets.

V. CONCLUSIONS

1. The use of an automated system to detect statistical relationships and statistically significant factor effects can reduce the number of erroneous results in evidence-based medicine, especially with widespread access to such a system that can be easily provided on the NI LabVIEW platform by connecting to the Internet.

2. The use of software environment NI LabVIEW is a promising area of building information processing and analysis systems. One of the important advantages is the ability to modify algorithms even by non-programmers.

3. One of the significant advantages of the NI LabVIEW platform is the simplicity of interaction with research equipment and other software systems.

This provides the basis for creating flexible automated workplaces.

4. Prospects for improving the system imply further addition of modules with the implementation of additional methods of statistical analysis, as well as providing easy access to the system by users, for example, via Internet.

REFERENCES

[1] Леонов В.П. Ошибки статистического анализа биомедицинских данных. / Международный журнал медицинской практики. – 2007, вып. 2, стр.19-35.

[2] Леонов В.П. Доказательная или сомнительная? Медицинская наука Кузбасса: статистические аспекты. – Томск: Томский государственный университет, 2010. – 175 с.

[3] Леонов В.П. Почему и как надо учить медиков статистике? (Доклад на международной конференции по доказательной медицине в Ереване 18-20.10.2012)

[4] Леонов В.П., Ижевский П.В. Об использовании прикладной статистики при подготовке диссертационных работ по медицинским и биологическим специальностям / Бюллетень ВАК, 1997, №5.

[5] LabVIEW for Everyone / Jeffrey Travis, Jim Kring. – Prentice Hall, 2006. – 1032 p.

[6] Програмування в NI LabVIEW. Технологія розробки віртуальних приладів : навч. посіб. / О.Г. Кисельова, А.В. Соломін. – К. : НТУУ “КПІ”, 2014. – 276 с.

[7] LabVIEW Fundamentals. – National Instruments Corporation, 2005. – 165 p.

[8] An introduction to statistical methods and data analysis / R.Lyman Ott, Michael Longnecker:5th ed. – USA: DUXBURY, 2001. – 1213 p.

[9] An Introduction to the Science of Statistics: From Theory to Implementation / Joseph C. Watkins: [Електронний ресурс]. – Режим доступу до ресурсу: https://www.math.arizona.edu/~jwatkins/statbook.pdf (дата звернення 28.02.2021)

[10] Introduction to Statistics. Online Edition / Primary author and editor David M. Lane: [Електронний ресурс]. – Режим доступу до ресурсу: https://onlinestatbook.com/Online_Statistics_Education.pdf (дата звернення 28.02.2021)

УДК 004.93

СИСТЕМА СТАТИСТИЧНОГО АНАЛІЗУ ДАНИХ ДЛЯ ДОКАЗОВОЇ МЕДИЦИНИ НА ПЛАТФОРМІ LabVIEW

Соломін А.В., доцент
andr-sol@i.ua
Сані Ахмед, студент
ahmed.shahahmedshah@hotmail.com
Факультет біомедичної інженерії
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"
м. Київ, Україна

Реферат *– Проблема створення інструментальних засобів для спрощення коректного використання статистичних методів досліджень залишається актуальною, особливо в біомедичній галузі, де фахівці не надто володіють математичним апаратом.*
В статті запропоновано реалізацію в програмному середовищі NI LabVIEW напівавтоматизованої системи, що забезпечує найчастіше використовувані засоби статистичного аналізу, але при цьому виконує перевірку необхідних критеріїв і меж їх використання, а також обирання відповідних статистичних методів в автоматичному режимі. Реалізація в програмному середовищі NI LabVIEW забезпечує легке підключення до апаратури, універсальність та простоту подальшої модифікації алгоритмів роботи.
Ключові слова: *статистичні методи, доказова медицина, надійність висновків, ймовірність помилок, критерії застосування, деревоподібна структура, дисперсійний аналіз, NI LabVIEW*

УДК 004.93

СИСТЕМА СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ ДЛЯ ДОКАЗАТЕЛЬНОЙ МЕДИЦИНИ НА ПЛАТФОРМЕ LabVIEW

Соломин А.В., доцент
andr-sol@i.ua
Сани Ахмед, студент
ahmed.shahahmedshah@hotmail.com
Факультет биомедицинской инженерии
Национальный технический университет Украины
"Киевский политехнический институт имени Игоря Сикорского"
г. Киев, Украина

Реферат *– Проблема создания инструментальных средств для упрощения корректного применения статистических методов исследований остается актуальной, особенно в биомедицинской отрасли, где специалисты не слишком хорошо владеют математическим аппаратом. В статье предложена реализация в программной среде NI LabVIEW полуавтоматической системы, обеспечивающей применение часто используемых методов статистического анализа, но при этом выполняющей проверку необходимых критериев и границ их применимости, а также выбор соответствующих статистических методов в автоматическом режиме. Реализация в программной среде NI LabVIEW обеспечивает легкое подключение к аппаратуре, универсальность и простоту дальнейших модификаций алгоритмов работы.*
Ключевые слова: *статистические методы, доказательная медицина, надежность выводов, вероятность ошибок, критерии применимости, древовидная структура, дисперсионный анализ, NI LabVIEW*