UDC 620.1.08

An alternative method for assessing the agreement between test results

A. Korobko[1,2], Yu. Kotova[2]

[1] L. Pogorilyy Ukrainian Scientific Research Institute of Forecasting and Testing of Machinery and Technologies for Agricultural Production, Kharkiv branch, Velika Panasivska Str, 236, 61040, Kharkiv, Ukraine
ak82andrey@gmail.com
[2] Kharkiv National Automobile and Highway University, Yaroslava Mudrogo Str, 25, 61002, Kharkiv, Ukraine

Abstract

The paper proposes a new way of assessing the agreement between measurement results during test quality assurance procedures in the laboratory. The decision-making rule is based on the measurement uncertainty. The probability, with which mathematical expectations of the measured data lie within the uncertainty of the indicator measurements, is proposed as a quantitative indicator. Such a quantitative indicator of the impact of methodological errors is proposed as the ratio of the difference between mathematical expectations of the measurement results obtained in different series of measurements to the average uncertainty of their determination, taking into account the applied decision-making rule. This indicator is based on the assumption that the measurement results are normally distributed. To simplify the process of calculating quality assurance indicators and reduce the risk of making mistakes in calculations, the "Agreement" module was created on the basis of a virtual test laboratory. The input data for the calculation of the agreement in the module are: measuring equipment, measurement results themselves (at least three for each tester), the coverage coefficient, full names of the people performing the measurements, and the coefficient of the decision-making rule.

Keywords: decision-making rule; measurement uncertainty; agreement between results; measurement; probability; random error; methodological error; error of determining the mean value.

Received: 25.01.2024 Edited: 27.02.2024 Approved for publication: 29.02.2024

Introduction

Testing of products at all stages of their life cycle is an important element in ensuring their quality. The test process also requires quality assurance and control (ensuring the reliability of results). This is required by the international standard EN ISO/IEC 17025:2017 [1]. In particular, paragraph 7.7.1 of this standard regulates that "*the laboratory shall have a procedure for monitoring the validity of results, the resulting data shall be recorded in such a way that trends are detectable and, where practicable, statistical techniques shall be applied to review the results...*" At the stage of ensuring the quality of test results, the task of assessing the accuracy and adequacy of measurements of indicators arises, with determining the degree of influence of random and methodological errors on the measurement result and establishing the criteria by which they should be evaluated. These tasks are quite complex and ambiguous.

The paper proposes a new method for assessing the agreement between measurement results when performing test quality assurance procedures in the laboratory. The decision-making rule is based on the uncertainty of measurement. The probability, with which mathematical expectations of the measured data lie within the uncertainty of the indicator measurements, is proposed as a quantitative indicator. Such a quantitative indicator of the influence of methodological errors is the ratio of the difference between mathematical expectations of the measurement results obtained in different series of measurements to the average uncertainty of their determination, taking into account the applied decision-making rule. This indicator is based on the assumption that the measurement results are normally distributed.

The main factors influencing test results in the testing laboratory are [1, 2]:

– operator (test subject, person or several people conducting the test);

– test equipment, measuring equipment (accuracy, calibration characteristics);

– environmental conditions (temperature, humidity, atmospheric pressure, air pollution, wind, etc.);

– time between individual measurements.

All of these factors are individual for each laboratory and are evaluated using the laboratory bias index [2].

In addition to the laboratory bias, in any testing laboratory, test results are influenced by random errors [2].

Traditionally, test quality assurance procedure is conducted on the basis of indicators regulated by the ISO 5725 standards [2]: reproducibility dispersion, repeatability variance, convergence limit, reproducibility limit. However, these standards do not regulate the limit values of these indicators, at which it is considered that test results are acceptable/unacceptable.

Certain developments to address this issue are reflected in the works of foreign authors. These works, as a rule, are of an applied nature – analytical tests and tests in medical laboratories.

The work [3] provides a brief overview of various types of external test quality assurance programmes. External quality assessment programmes were initiated in the middle of the 20th century, indicating wide discrepancies between laboratory results. They have been developed in different countries and some mechanisms for their harmonization have been proposed: establishing common performance specifications, having quality assurance programmes as a training tool. It is emphasized that participation in external quality assessment programmes allows one to know the real uncertainty of test results and whether the results of an individual laboratory are consistent with those of other laboratories that use the same analytical method.

The work [4] noted that technical quality assurance and quality control are important activities in medical laboratories to ensure the appropriate quality of the test results obtained. The proposed quality control framework provides methods and tools by which the performance of different tools can be compared in a so-called error detection table to enable optimization and verification of the overall test plan.

However, this work does not provide a specific mathematical apparatus in terms of alarm frequency, as well as pre-analytical, analytical and post-analytical error detection performance.

Specific methods for calculating reproducibility and repeatability characteristics are given in [5–8]. However, these works are of an applied nature and do not provide general guidelines for standardizing reproducibility and repeatability indicators in the laboratory.

A significant number of works, both fundamental textbooks on probability theory and experimental theory, and publications on solving particular problems [9–12], are directly devoted to the issue of studying the agreement between the distributions of experimental data. However, it should be noted that to our knowledge until today, there is no unified approach to terms and definitions on this issue found in the literature. In this regard, discrepancies arise in the interpretation of individual concepts.

A detailed analysis of indicators and criteria for establishing compliance with the theoretical distribution was made experimentally in [13, 14]. These indicators are called "agreement criteria". These are the criteria of Pearson $\chi 2$, Kolmogorov – Smirnov, Cramér – von Mises – Smirnov, Anderson – Darlign, Cooper and others. Also, to check the adequacy of a mathematical model to experimental data, a criterion is used based on a comparison of the reproducibility variance of the mean value of the response function and the variance of adequacy [15, 16].

The works [17, 18] present a mathematical apparatus for assessing the compliance of a theoretical model with experimental measurements, used in the development of new test methods. However, to formulate a decision-making rule when ensuring the quality of tests, the specified mathematical apparatus requires certain clarifications.

The works [19, 20] emphasize that measurement uncertainty is an effective tool for developing decision-making rules both in laboratory activities and in studying the quality of industrial processes.

Each of the criteria under consideration has its own scope of application. However, as a general disadvantage, one can highlight the fact that they require massive sampling (more than 50, and in rare cases, more than 300 observations). Existing criteria do not fully answer the question of how quantitatively the theoretical and experimental data agree. Another disadvantage is that the existing criteria take into account the metrological characteristics of measurements integrally, i.e. when assessing the agreement between the distributions, the uncertainty or error, with which the measurements are performed, is not considered separately.

In addition, the problem of determining the limit value of the criterion indicator at which a decision is made on the consistency/inconsistency of the test results is unresolved.

The purpose of the study is to substantiate the decision-making rule when ensuring the quality of tests in the laboratory based on measurement uncertainty. To achieve these goals, it is necessary to solve the following tasks:

– justify the indicator and criterion for establishing a decision-making rule based on the results of ensuring the quality of tests in the laboratory using measurement uncertainty;

– examine the proposed indicator of agreement;

– develop structural information for software calculation of quality assurance indicators during laboratory testing.

Research results

Indicator and criterion for establishing a decision-making rule based on results of quality assurance of laboratory tests using measurement uncertainty.

Test methods are implemented by measuring the values of certain parameters using a specific method and regulated measuring equipment. Of course, both procedure and measuring instruments are not free of errors. In this regard, the results of experimental studies obtained under conditions of reproducibility or repeatability [2] will differ from each other. This is due to the influence of random, systematic and methodological errors.

The development of computer technologies has made it possible to increase the accuracy of obtaining experimental measurement data by automating their acquisition and processing. Let us assume that studies of one parameter are carried out using different methods (by different testers, under different conditions). We shall roughly call them as Method 1 and Method 2. In the future, the numbers 1 and 2 in the formula indices will indicate the results obtained using Method 1 and Method 2, respectively.

Considering the array of values of quantities obtained by Method 1 and the array of values obtained by Method 2 as two independent random variables, it is necessary to determine the probability of their joint occurrence, taking into account the value of the intervals in which the actual value of the measured quantity may be located, which will be a criterion for the comparability of results.

The measurement model (model equation) of the indicator under consideration has the following form:

$$y = f\left(X \pm U_X\right), \tag{1}$$

where $X = \{x_1, x_2, \ldots, x_n\}$ is a set of values of quantities included in the model equation for measuring the theoretical value of the considered y indicator;
$U_X = \{U_{x1}, U_{x2}, \ldots, U_{xn}\}$ is a set of values of expanded uncertainties with which the X indicators are measured and which are included in the model equation for measuring the considered y indicator;
n is a number of quantities included in the model equation for measuring the considered y indicator.

Let us assume that the measurement results are distributed according to the normal law.

The uncertainty of the U_X measurement is calculated with a confidence probability of $P = 0.95$, so the limits of the change in the y indicator will cover almost the entire area under the distribution curve. Based on the definition of the concept of "measurement uncertainty" and (1), it can be stated that the rough limits of the possible values of the y indicator will be determined by the magnitude of its measurement uncertainty

$$y_{\max}\left(y_{\min}\right) = \bar{y} \pm U_y, \tag{2}$$

where $y_{\max}$, $y_{\min}$ are possible maximum and minimum values of y, respectively;
U_y is the uncertainty of the measurement of the y indicator.

Regardless of the method of determining the y indicator, its mathematical expectation, standard deviation, and error in determining the mean value of experimental data are determined by classical formulas of probability theory [15]:

$$\bar{y} = \frac{\sum_{i=1}^{j} y_i}{j}; \tag{3}$$

$$\sigma_y = \sqrt{\frac{1}{j-1}\sum_{i=1}^{j}\left(y_i - \bar{y}\right)^2}; \tag{4}$$

$$m_y = \frac{\sigma_y}{\sqrt{j}}, \tag{5}$$

where y_i is the i-th value of the measured y quantity (indicator);
j is the number of measurements of the y value.

In real conditions, due to the impact of random and systematic errors, the uncertainty of measuring the y indicator by Method 1 and Method 2 will differ:

$$U_{y1} \neq U_{y2}. \tag{6}$$

The mathematical expectations of the y indicator will also differ:

$$\bar{y}_1 \neq \bar{y}_2. \tag{7}$$

As an indicator for assessing the agreement between measurement results obtained by Method 1 and Method 2, an indicator is proposed, which is determined by the ratio of the difference in mean values (Δy) to the combined uncertainty of their measurements ($U_{y\Sigma}$), taking into account the applicable (adopted) decision-making rule r

$$k_\text{c} = \frac{\Delta y}{U_{y\Sigma}}, \tag{8}$$

where

$$\Delta y = \left|\bar{y}_1 - \bar{y}_2\right|; \tag{9}$$

$$U_{y\Sigma} = r\sqrt{U_{y1}^2 + U_{y2}^2}, \tag{10}$$

r is the decision-making rule.

Fig. 1 is a graphical interpretation of agreement between the measurement results obtained by Method 1 and Method 2.

Using special tables given in the literature on mathematical statistics (for example, [15]), depending on the value of k_c, the probability, with which the mathematical expectation of a value determined by one of the methods goes beyond the uncertainty of

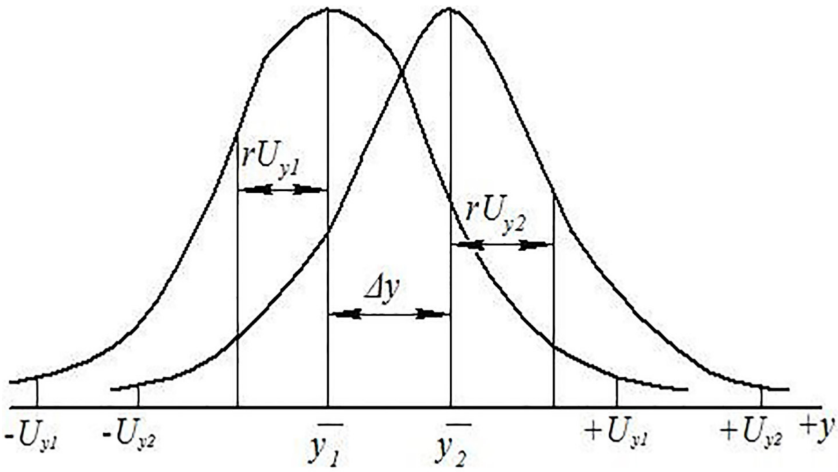


Fig. 1. Graphical interpretation of agreement between the measurement results obtained by Method 1 and Method 2

Table 1

A brief table of probability values for agreement between distributions

k_c	0.05	0.1	0.15	0.2	0.3	0.4	0.6	0.8	1.0	1.2
P_c	0.9799	0.9203	0.8985	0.8415	0.7642	0.6892	0.5485	0.4237	0.3173	0.2301

measurement of the same value determined by another method, taking into account the applied method of decision-making rule, is determined. From the above definition we can conclude that the k_c indicator is a quantitative indicator of the expression of the total error of the measurement method used.

To determine the probability of agreement (P_c) between measurement results obtained by different methods (taking into account the applied decision-making rule), it is necessary to find the value

$$P_c = 1 - P_{k_м}, \tag{11}$$

where $P_{k_м}$ is the probability of discrepancy between theoretical and experimental data [15].

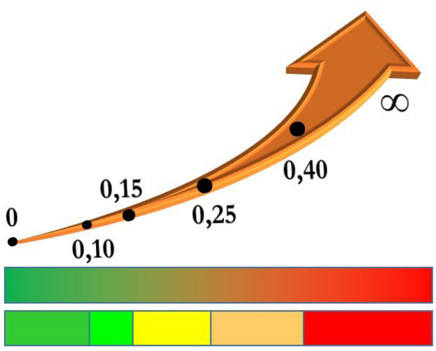
Study of the agreement indicator k_c

The value of the k_c indicator varies within the limits k_c = [0; ∞). Table 1 shows the values of the k_c indicator and their corresponding probabilities of agreement between distributions.

As can be seen from Table 1, with a value of k_c = 0.6, the probability of agreement between two distributions will be 54%. That is, with a value of ks≈0.68, the probability of agreement between two distributions will be 50%. With a value of ks⩽0.25, the probability of agreement between the results will be 80% or more. In this case, we can assume that the methodological error has little effect on measurement results, and the results themselves are comparable (the quality of testing in the laboratory is ensured).

The value of the k_c indicator significantly depends on the value of the combined measurement uncertainty $U_{yΣ}$. The larger the value of $U_{yΣ}$ is, the more likely the mean values of measurement distributions will be within the indicator measurement uncertainty. Therefore, the range of measurement uncertainty limits is limited by the laboratory decision-making rule, r (see formula 10). This can be any number in the range (0; 1].

Fig. 2 shows the proposed gradation for assessing the degree of agreement between measurement results during comparative tests based on the k_c indicator:



(0.40; ∞] – test results are incomparable;

(0.25; 0.40] – the agreement between the results is unsatisfactory;

(0.15; 0.25] – the agreement between the results is satisfactory;

(0.10; 0.15] – the agreement between the results is good;

[0; 0.10] – the agreement between the results is very good.

Fig. 2. Gradation of assessment of the degree of agreement between measurement results during comparative tests based on k_c indicator

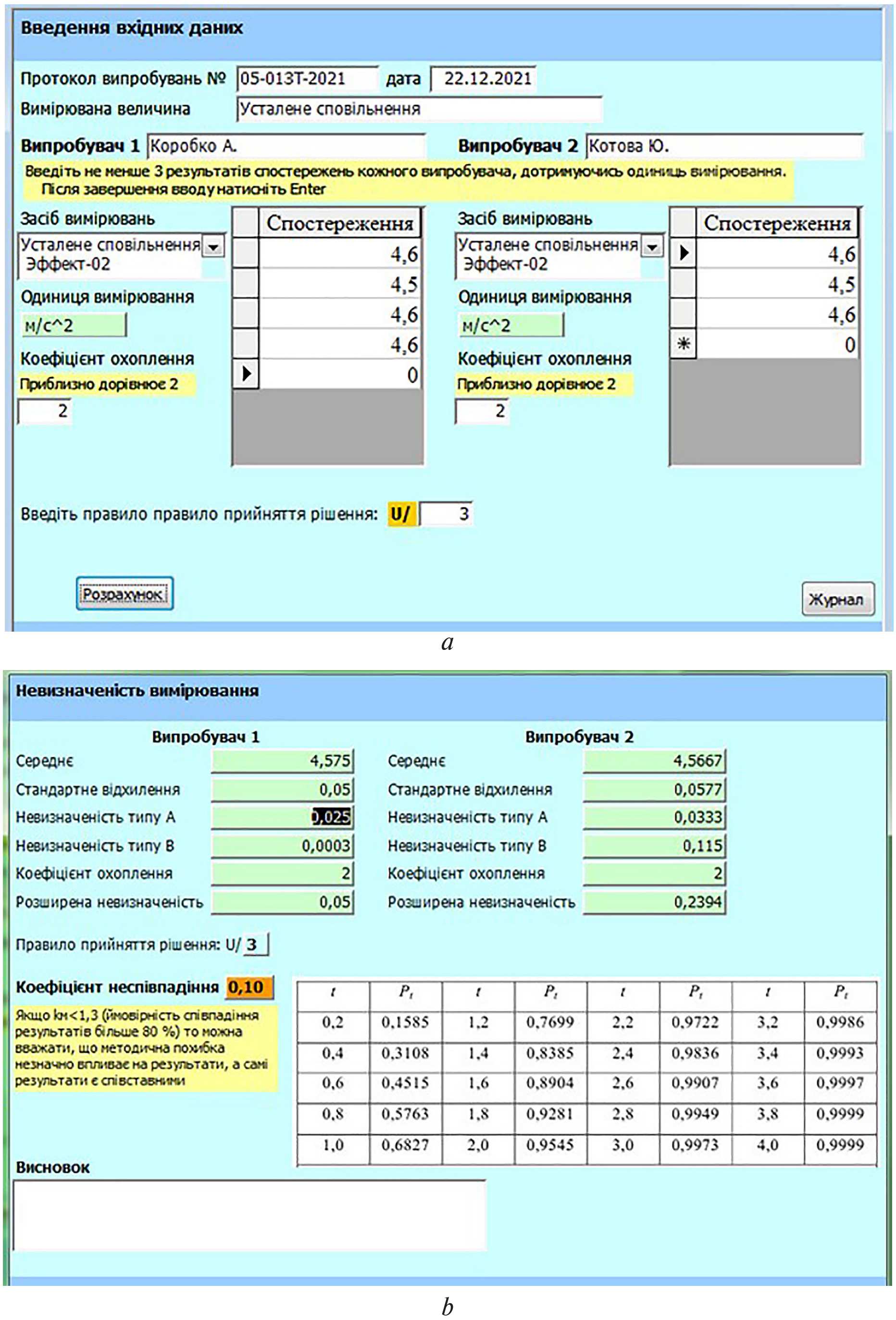


Fig. 3. Agreement Module Interface:
a – data entry window; *b* – calculation results window

The data structure of the software calculation of quality assurance indicators during testing in the laboratory

To simplify the process of calculating the quality assurance indicators and reduce the risks of errors in calculations, it is proposed to create an "Agreement" module on the basis of a virtual testing laboratory [21]. The interface of this module is shown in Fig. 3.

Structurally, the Agreement module consists of:

- tables: "Report", "Measurement results – Tester 1", "Measurement results – Tester 2", "Type of Measurements"; moreover, these tables are connected by a "One-to-Many" ratio in the "Report Code" field with data integrity and cascading updates;
- queries: "Statistics", "Uncertainty", "Uncertainty A", "Uncertainty A_STDev", "Uncertainty B", "Agreement Report", which provide the calculation of the agreement between the results;
- report "Comparative Test Report", which is prepared in the format that the laboratory provides. In addition, a journal for calculating the measurement uncertainty is prepared separately;
- forms: "Data entry", "Result".

The input data for the calculation of the agreement in the module are: measuring equipment, the measurement results themselves (at least three for each tester), the coverage coefficient, full names of the people performing the measurements and the coefficient of the decision-making rule. After pressing the "Calculation" button, all the information necessary to interpret the results of comparative tests is displayed on the screen. Pressing the "Report" button generates and displays on the screen a comparative test report for in the provided format, which can be printed or saved in PDF format.

Conclusions

As a result of the study, a new method for assessing the agreement between measurement results during test quality assurance procedures in the laboratory has been developed. The decision-making rule is based on the uncertainty of measurement. The probability, with which mathematical expectations of the measured data lie within the uncertainty of the indicator measurements, is proposed as a quantitative indicator. Such a quantitative indicator of the influence of methodological errors is proposed as the ratio of the difference between mathematical expectations of the measurement results obtained in different series of measurements to the average uncertainty of their determination, taking into account the applied decision-making rule. This indicator is based on the assumption that the measurement results are normally distributed.

Альтернативний показник оцінювання збігу результатів випробувань

А.І. Коробко[1,2], Ю.М. Котова[2]

[1] Харківська філія Українського науково-дослідного інституту прогнозування та випробування техніки і технологій для сільськогосподарського виробництва імені Леоніда Погорілого, вул. Велика Панасівська, 236, 61040, Харків, Україна
ak82andrey@gmail.com
[2] Харківський національний автомобільно-дорожній університет, вул. Ярослава Мудрого, 25, 61002, Харків, Україна

Анотація

Метою дослідження є обґрунтування правила прийняття рішення при забезпеченні якості випробувань у лабораторії на основі невизначеності вимірювань. Для досягнення мети вирішено такі завдання: обґрунтовано показник і його значення критеріїв для встановлення правила прийняття рішення за результатами забезпечення якості випробувань у лабораторії із застосуванням невизначеності вимірювання; досліджено запропонований показник збігу; розроблено структуру даних програмного розрахунку показників забезпечення якості при випробуваннях у лабораторії. Правило прийняття рішення ґрунтується на основі невизначеності вимірювання. Як кількісний показник запропоновано ймовірність, із якою математичні очікування виміряних даних знаходяться у межах невизначеності вимірювання показника. Таким кількісним показником впливу методичної похибки вважається відношення різниці між математичними очікуваннями результатів, отриманих у різних серіях вимірювань, до середньої невизначеності їхнього визначення з урахуванням застосованого правила прийняття рішення. Показник ґрунтується на припущенні, що результати вимірювань нормально розподілені. Значення показника збігу суттєво залежить від значення спільної невизначеності вимірювання. Тому інтервал меж невизначеності вимірювання обмежується правилом прийняття рішення у лабораторії. Для спрощення процесу розрахунку показників забезпечення якості та зменшення ризиків допущення помилок у розрахунках на базі віртуальної випробувальної лабораторії створено модуль "Agreement". Вхідними даними для розрахунку збігу в модулі є засоби вимірювання, безпосередньо результати вимірювань (не менше трьох у кожного випробувача), коефіцієнт охоплення, прізвища осіб, що здійснюють вимірювання, та коефіцієнт правила прийняття рішення.

Ключові слова: правило прийняття рішення; невизначеність вимірювання; збіг результатів; вимірювання; ймовірність; випадкова похибка; методична похибка; похибка визначення середнього значення.

References

1. DSTU EN ISO/IEC 17025:2019. General requirements for the competence of testing and calibration laboratories (EN ISO/IEC 17025:2017, IDT) (in Ukrainian).
2. ISO 5725. Accuracy (trueness and precision) of measurement methods and results (6 parts). Available at: https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en
3. Carmen Ricós, Pilar Fernández-Calle, Carmen Perich, Sverre Sandberg. External quality control in laboratory medicine. Progresses and future. *Advances in Laboratory Medicine*, 2022, vol. 3, issue 3, pp. 221–231. doi: 10.1515/almed-2022-0058
4. Huub H. van Rossum. Technical quality assurance and quality control for medical laboratories: a review and proposal of a new concept to obtain integrated and validated QA/QC plans. *Critical Reviews in Clinical Laboratory Sciences*, 2022, vol. 59(8), pp. 586–600. doi: 10.1080/10408363.2022.2088685
5. Measurement Process Characterization. Gauge R&R studies. Available at: https://goo.su/GDSFtnn
6. Maha Al-Eid, Mohamed M. Shoukri. On the Index of Repeatability: Estimation and Sample Size Requirements. *Open Journal of Statistics*, 2019, vol. 9, no. 4, pp. 530–541. doi: 10.4236/ojs.2019.94035
7. Eduarda Filipe. Validation of calibration methods – a practical approach. *Advanced Mathematical and Computational Tools in Metrology VI*, 2004, pp. 285–292. doi: 10.1142/9789812702647_0027
8. Bartlett J.W., Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics & Gynecology*, 2008, vol. 31, issue 4, pp. 466–475. doi: 10.1002/uog.5256
9. Dieter Rasch, Rob Verdooren, Jürgen Pilz. Applied Statistics: Theory and Problem Solutions with R. Wiley, 2019. 512 p.
10. Jeff Wu C.F., Hamada M. Experiments: Planning, Analysis, and Optimization. John Wiley & Sons, 2021. doi: 10.1002/9781119470007
11. Steven A. Weissman, Neal G. Anderson. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Organic Process Research & Development*, 2014, vol. 19(11). doi: 10.1021/op500169m
12. Quality and Reliability of Technical Systems – Theory, Practice, Management (2nd ed.). *Assembly Automation*, 1998, vol. 18, no. 4. doi: 10.1108/aa.1998.03318dad.012
13. Frank Emmert-Streib, Matthias Dehmer. Understanding Statistical Hypothesis Testing: The Logic of Statistical Inference. *Machine Learning Knowledge Extraction*, 2019, vol. 1(3), pp. 945–961. doi: https://doi.org/10.3390/make1030054
14. Konold T.R., Fan X. Hypothesis Testing and Confidence Intervals International. Encyclopedia of Education (3rd ed.), 2010, pp. 216–222. doi: 10.1016/B978-0-08-044894-7.01337-3
15. Bhattacharya P.K., Burman Prabir. Probability Theory. Theory and Methods of Statistics, 2016, pp. 1–24. doi: 10.1016/B978-0-12-802440-9.00001-1
16. Zulfiqar Ali, S Bala Bhaskar. Basic statistical tools in research and data analysis. *Indian J Anaesth.*, 2016, vol. 60(9), pp. 662–669. doi: 10.4103/0019-5049.190623
17. Korobko A. Measurement Uncertainty as a Test Model Assessment Tool. *Advanced Optoelectronics and Lasers CAOL-2019: Proceedings of 2019 IEEE 8th International Conference with XVI Scientific Workshop "Measurement Uncertainty: Scientific, Normative, Applied and Methodical Aspects" (UM-2019)*. Sozopol, Bulgaria, 06–08 September 2019, pp. 707–710. doi: 978-1-7281-1814-7/19/31:00
18. Podrigalo M., Verbitskiy V., Korobko A., Baidala V. Otsiniuvannia adekvatnosti rezultativ doslidzhen, shcho otrymani riznymy metodamy (v poriadku obhovorennia) [Researches obtained by different methods and their assessment of adequacy (in order of discussion)]. *Perspective Technologies and Devices*, 2019, vol. 15, pp. 63–67 (in Ukrainian). doi: https://doi.org/10.36910/6775-2313-5352-2019-15-9
19. Volodarsky E.T., Kosheva L.O., Klevtsova M.O. Approaches to the Evaluation of Conformity Taking into Account the Uncertainty of the Value of the Monitored Parameter. *Advanced Optoelectronics and Lasers CAOL-2019: Proceedings of 2019 IEEE. 8th International Conference with XVI Scientific Workshop "Measurement Uncertainty: Scientific, Normative, Applied and Methodical Aspects" (UM-2019)*. Sozopol, Bulgaria, 06–08 September 2019, pp. 648–652. doi: 978-1-7281-1814-7/19/31.00
20. Volodarsky E., Kosheva L., Klevtsova M. Formation of the rule decision-making about suitability products on the basis of the adaptive algorithm. *Proceedings of 28th International Scientific Symposium "Metrology And Metrology Assurance"*, 2018, pp. 331–335.
21. Korobko A., Shatikhina V. Virtualnyi trenazher akredytovanoi vyprobuvalnoi laboratorii [Virtual simulator for test laboratory accredited]. *Perspective Technologies and Devices*, 2020, vol. 17, pp. 72–78 (in Ukrainian). doi: https://doi.org/10.36910/6775-2313-5352-2020-17-11