UDC 616-093 DOI:

Ensuring the validity of the results of calibrations and tests by statistical methods

A. Kolbasin [1,2]

[1] *State Enterprise "Kharkiv Regional Research and Production Center for Standardization, Metrology and Certification", Myronosytska Str., 36, 61002, Kharkiv, Ukraine*
[2] *National Scientific Centre "Institute of Metrology", Myronosytska Str., 42, 61002, Kharkiv, Ukraine*
kai_2548@i.ua

Abstract

According to the requirements of ISO/IEC 17025:2017, the validity of test and calibration results is ensured, inter alia, by intralaboratory check of the results obtained. In this case, it is preferable to use statistical methods.

The ISO 5725 standards define a number of such methods, but the choice of specific methods is left to the laboratory, taking into account the requirements for the adequacy of the effort, resources and time for the purposes of the work performed and the risks of obtaining inappropriate results. In this case, the laboratory itself must in a certain way determine which objects of calibrations (tests) should be predominantly used in checks and what frequency of checks should be foreseen.

In connection with the increase in the accuracy and complexity of measuring systems, the need to apply the methods of the theory of random processes becomes more and more obvious. It is shown that the use of the Poincaré plot makes it possible to comprehensively, effectively and visually evaluate changes in the measuring process from the point of view of the dynamics of the obtained measurement results.

The results of the check, in particular, the intermediate precision, make it possible to obtain a more realistic evaluation of measurement uncertainty in accordance with ISO 21748.

The paper analyses some practical approaches (of varying degrees of complexity) to intralaboratory checks of the validity of calibration (test) results.

Keywords: calibration; tests; intralaboratory check; Poincaré plot; ANOVA.

Received: 07.10.2021 Edited: 12.11.2021 Approved for publication: 18.11.2021

1. Introduction

Standard [1] establishes requirements for testing and calibration laboratories for the ability to obtain valid results.

At the same time, according to the requirements of [1, 2], the validity of test and calibration results is ensured, inter alia, by intralaboratory checks of the results obtained. In this regard, the use of statistical methods has advantages in identifying trends and assessing the significance of possible deviations when varying factors that affect the measurement result, but cannot be completely controlled.

In the ISO 5725 series [3], a number of such methods is defined, however, the choice of specific methods remains with the laboratory, taking into account the requirements for the adequacy of the effort, resources and time for the purposes of the work performed and the risks of obtaining inappropriate results. In addition, the laboratory must itself determine which calibration (test) objects should be predominantly used in checks and what frequency of checks should be envisaged.

Within the framework of a single laboratory, it is most often limited to the determination of repeatability and intermediate precision, taking into account such factors as "operator", "time interval between measurements", less often "environmental parameters" and "calibrations".

Intralaboratory precision values can serve as a refinement [4] to the GUM model approach [5] in evaluation of the measurement uncertainty.

The paper analyzes practical approaches to intralaboratory checks of the validity of the results of calibrations (tests).

2. Statement of the main material

The main purpose of an intralaboratory check is to ensure that variations in the factors "operator" and "time interval between measurements" do not lead to a significant increase in either the uncertainty declared by the laboratory in its measurement and calibration capabilities, or the target uncertainty.

Of course, it is necessary to determine which item or several items of calibration (test) will be used during the check. Most often, any method is intended for a certain category of devices, which may differ both in measurement ranges and in the maximum permissible error. This creates a multi-parameter problem of

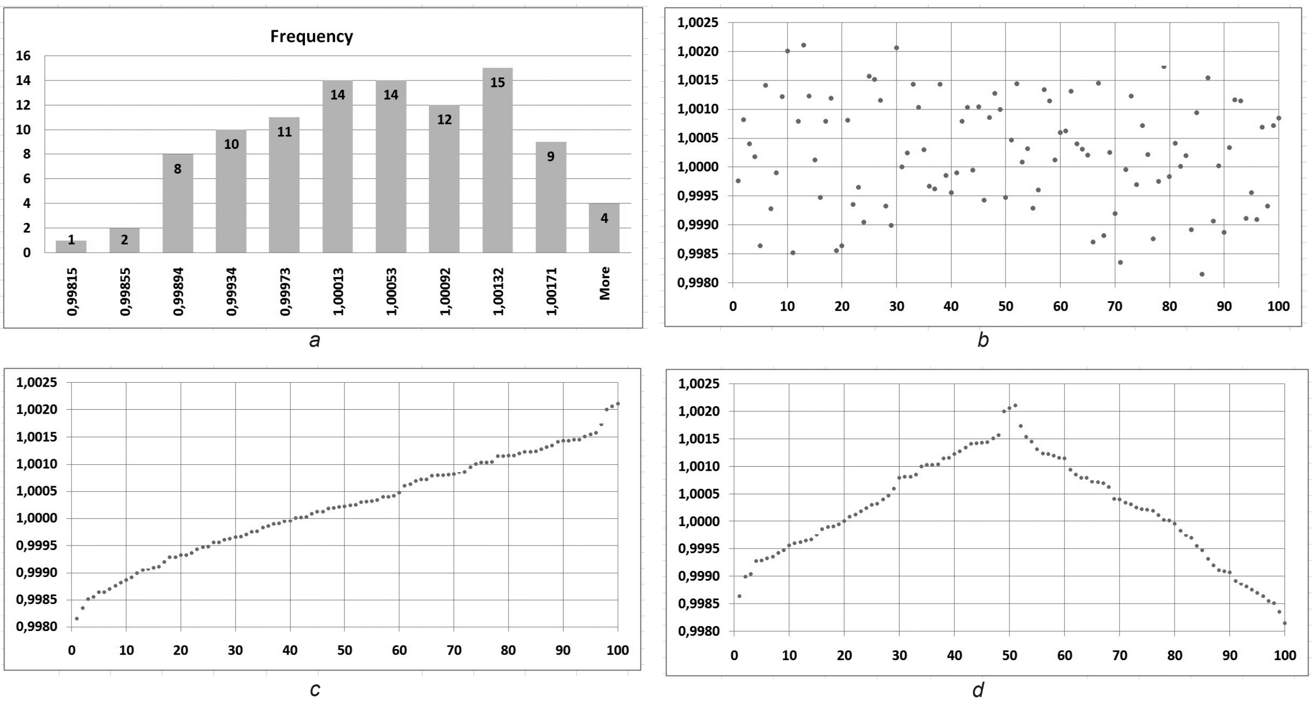

Fig. 1. A dataset of 100 normally distributed numbers: (*a*) – histogram; (*b*) – original dataset; (*c*) – the dataset has been sorted in ascending order; (*d*) – the dataset has been sorted partially in ascending order and partially in descending one

choosing a specific device(s) and measurement point(s), because the degree of influence of factors, on the one hand, may depend on the point of the measurement range, on the other, on the ratio of the inaccuracy caused by the influencing factor to the maximum permissible error of the device. The solution of this problem is generally entrusted to laboratory specialists.

Another task is to determine the frequency of such checks. As a rule, under stable working conditions of the laboratory and with constant staff, the frequency of checks can be from one to two years.

According to [3], the main indicators determined during intralaboratory control are repeatability and intermediate precision. In this case, at least two operators carry out a series of measurements at different times. The obtained data are processed by statistical methods.

It should be noted that there are some problems associated with establishing both the number of measurements in series and the number of series. If we proceed from the hypothesis that the measurement results are only random numbers, then it is desirable to increase the amount of data received. So, to determine the type of distribution law, it is recommended that the number of measurements in a series be no less than 15–30. But it is known that an increase in the number of measurements, as a rule, leads to the fact that starting from a certain number, the variance of the results can begin to increase [6]. This is especially noticeable when making accurate measurements using complex measuring systems.

This is due to both the drift of the measuring systems (instruments) and the change in the measured value over time. The reasons for the drift can be an unstable temperature regime, transient processes in the measuring circuit, instability of the characteristics of the devices used in the experiment, and the like.

As a result, when analyzing data only as random numbers, important information about the nature of changes in results over time may be lost [7, 8]. So, for example, the same histogram (Fig. 1(*a*)) describes a certain set of random numbers, regardless of the sequence of their appearance (Fig. 1(*b*) – Fig. 1(*d*)). Of course, the measurement results, as in Fig. 1(*c*), are unlikely to indicate a high quality of measurements.

In this case, analysis of the data, for example using control cards, can provide additional objective information about the measurement process.

It may also be useful to use the Poincaré plot, by the form of which it is possible to qualitatively characterize the measuring process from the point of view of its behavior as a kind of dynamical system. This tool is especially convenient when analyzing modern measuring systems, when the amount of information can reach several hundred or thousands of readings.

A Poincaré plot is a two-dimensional graphical representation of a time series where the ordinate is the n-th term in the sequence and the abscissa is the previous $(n-1)$-th term.

Fig. 2 shows Poincaré plots for some datasets representing some varieties of random processes.

Fig. 3 shows the experimental results of measuring the resistance ratio of two MC3006 measures (value 100 Ohm, accuracy class 0.001) using the comparator bridge MI AccuBridge. The graphs show the relative changes in the measured resistance ratios when the thermostat is off and on. The time interval between successive measurements is estimated to be of the order of 20 s.

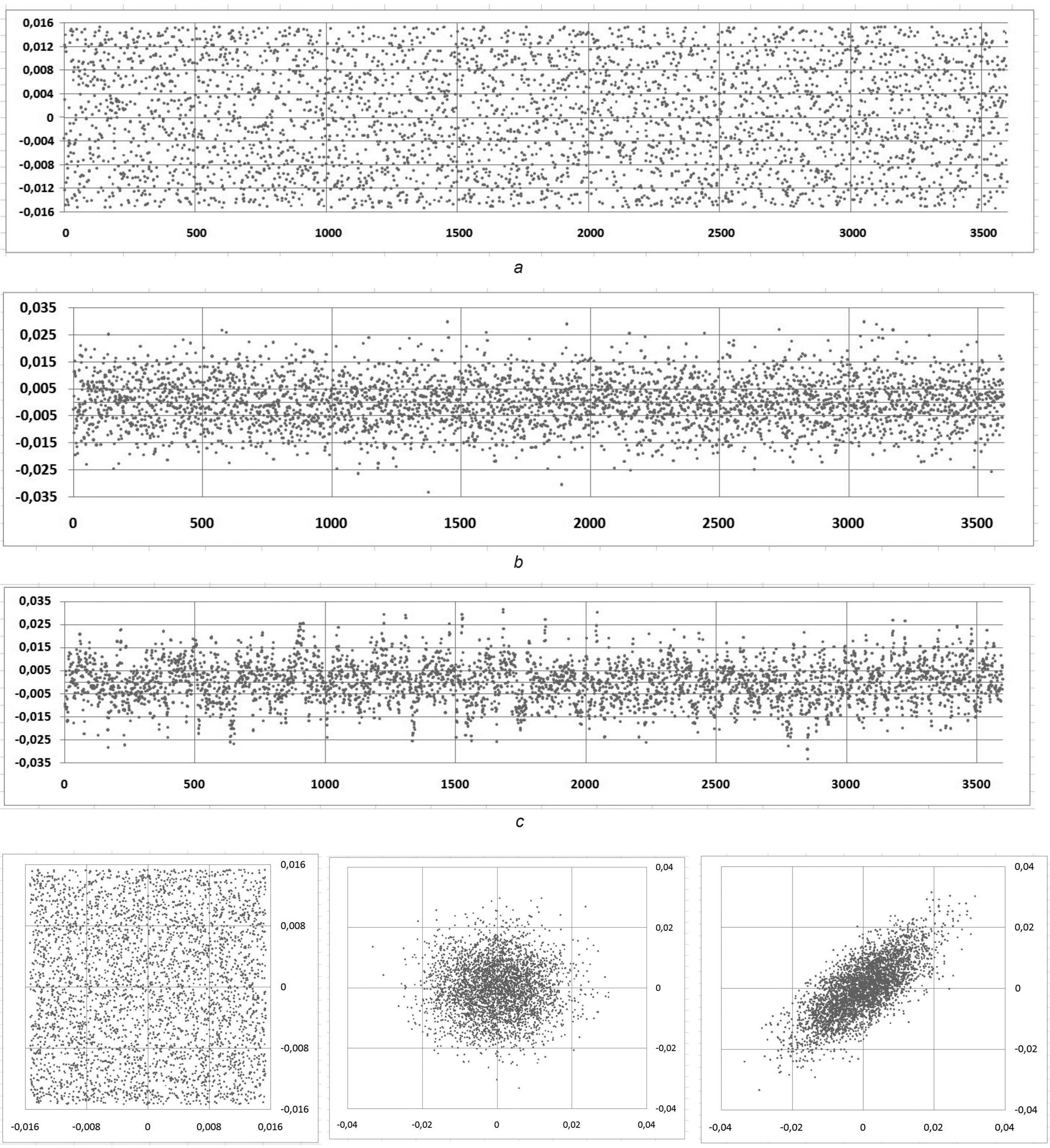

Fig. 2. Random processes with zero mean and equal variance, obtained by modeling, and the corresponding Poincaré plots: (*a*), (*d*) – "white noise" process with a uniform distribution of values; (*b*), (*e*) – Gaussian process of the "white noise" type; (*c*), (*f*) – Gaussian random process with exponential correlation function.

Comparison of Fig. 3(*c*) and Fig. 3(*d*) shows that when the thermostat is on, the random scatter of the bridge readings is significantly reduced: the relative deviations are in the range (-2.9×10^{-8}; 2.6×10^{-8}). The standard deviation of a single measurement of the ratio is estimated at 8.9×10^{-9}.

As a result, the Poincaré plot shown in Fig. 3(*d*), can be considered a "reference portrait" of the system under consideration, which can later be used for visual express analysis of changes during periodic diagnostics of the measuring system.

There are many programs and software packages, such as Microsoft Office Excel, R, SAS, Statistica, SPSS, MatLab and others, for statistical data processing.

It is recommended to analyse the data in the following sequence:

1) Detection of outliers.

To detect outlying variances are used Cochran's test (clause 8.3.4 of ISO 5725–2), if the number of elements in the series is the same, or Levene's test if the number of elements is different. Bartlett's test, Irwin's test, Lvovskiy's test, etc. are also used. Some

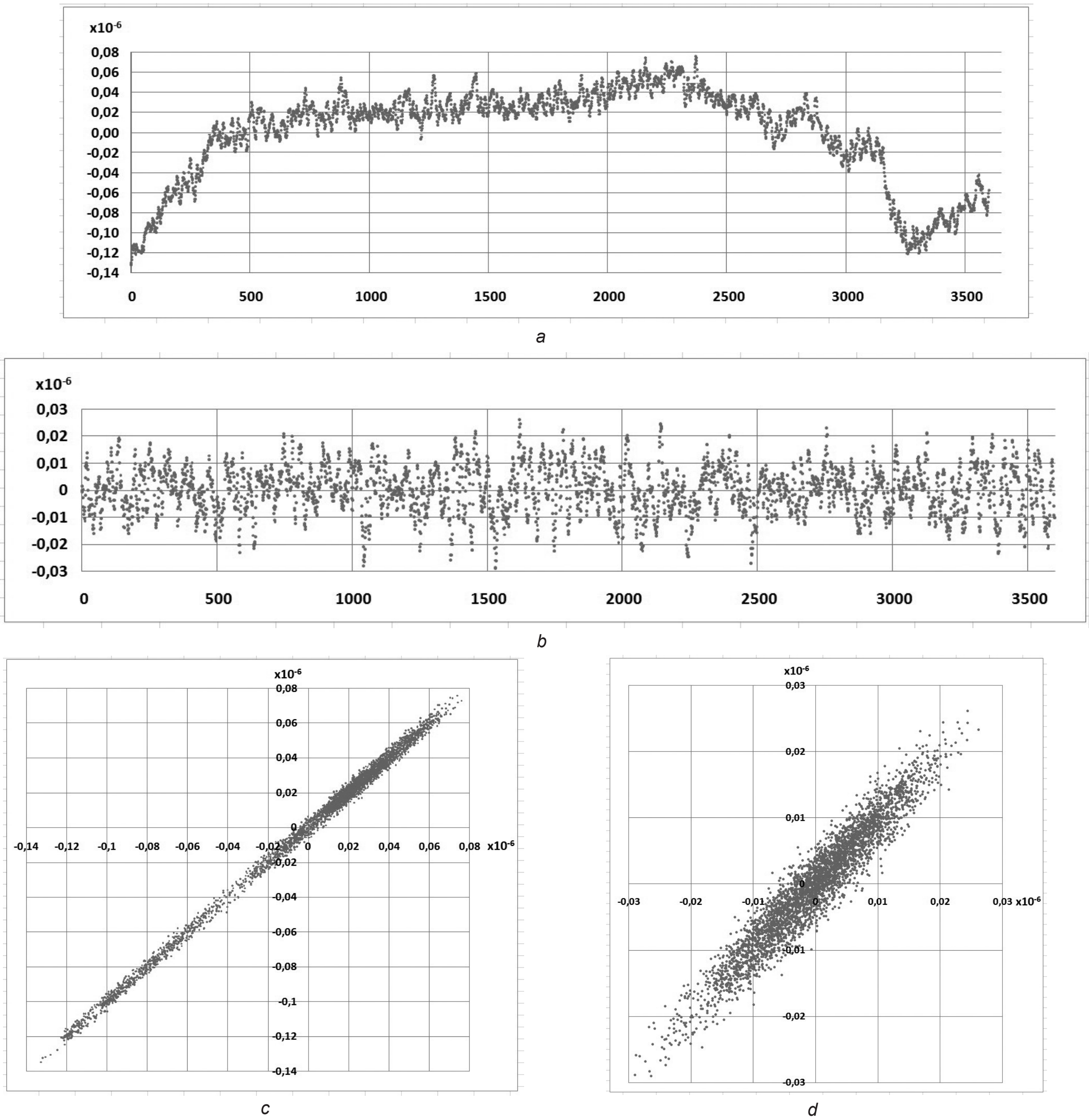

Fig. 3. Time series of measurement results and the corresponding Poincaré plots:
(*a*), (*c*) – with the thermostat off; (*b*), (*d*) – when the thermostat is on

methods for identifying and processing data outliers are also given in the standard [9].

If the Cochran's test statistic exceeds the 5% critical value, the series with the maximum variance is removed from the results. At the same time, the allowable number of seizures should be determined, exceeding which it is necessary to conduct additional research to establish the possible reasons for such an excessive amount of outliers.

The critical values for Cochran's test C_{crit} are given in many mathematical and statistical reference books. They can also be calculated using an Excel spreadsheet (Fig. 4).

If the check by Cochran's test has raised the suspicion that the high variation is due to only one of the measurement results, then the data are analyzed for possible outliers by the Grubb's test (clause 8.3.5 of ISO 5725–2).

2) Detection of data drift

Detection of drift (trend) can be carried out using the *t*-criterion (the method of differences in average levels), Foster-Stuart test (except for the trend itself, the latter allows to establish the presence of a dispersion trend), etc.

If the hypothesis of the randomness of the trend is not confirmed, then it is advisable to correct the initial data by the value of the trend (mainly linear).

It should be borne in mind that the presence of a trend may also indicate insufficient time to enter the steady state or insufficient observation time for

	A	B	C
1	significance level (5% or 1%)	5%	5%
2	number of compared series (p = number of laboratories at a given level - ISO 5725-2)	10	10
3	number of measurements in a series (n = number of test results per cel - ISO 5725-2)	5	5
4	FINV(probability,deg_freedom1,deg_freedom2) =	4.4552	=FINV(C1/C2;C3-1;(C2-1)*(C3-1))
5	C_{crit} =	0.3311	= C4 / (C4 + C2 - 1)

Fig. 4. Fragment of an Excel sheet with the calculation of Cochran's statistics C_{crit}

averaging ultra-low-frequency fluctuations in the measuring circuit.

3) Analysis of data on their belonging to the normal distribution law.

Compliance with the normal law can be established graphically; by the Pearson consistency criterion; Shapiro-Wilk test by Kolmogorov's test (application of Bolshev's correction to Kolmogorov test allows to reduce the required sample size), Romanovsky, Anderson-Darling, Cramer-von Mises-Smirnov, Chauvin, etc., as well as by indicators of asymmetry and kurtosis. Methods and criteria for checking the deviation of the probability distribution from the normal are also given in the standard [10].

In the case of a positive answer, it is advisable to use parametric criteria in further analysis, otherwise – nonparametric (robust) criteria.

4) Checking acceptability of measurement results obtained under conditions of intermediate precision.

The intermediate precision is determined in terms of two factors: "time" and "operator". A few operators perform multiple series of calibration measurements on different days. In this case, the average time interval between successive measurements and the sequence of changing operators should be established by the laboratory, based on the characteristics of the calibration procedure used and the conditions for performing measurements.

Testing the variance of the results in each series for homogeneity according to Fisher's test is easy to check using analysis of variance (ANOVA) [11].

Fig. 5 shows the results of processing using the Excel analysis package of four series of 10 measurements obtained by two operators in two days of measurements. Since the calculated value of the F statistic is less than the critical value F_{crit}, then at a significance level of $\alpha = 0.05$, the influence of factors in this case is statistically insignificant.

If it is necessary to assess the influence of each of the factors on the measurement accuracy, the measurement results should be subjected to a two- or three-factor ANOVA.

Results obtained with ANOVA (Excel) can be expressed as variances entered in [3]:

- repeatability variance – $s_r^2 = MS_W$, where MS_W – variance within groups;
- variance due to various factors:

$$s_B^2 = \begin{cases} (MS_B - MS_W)/\ Count & if \quad MS_W \le MS_B \\ 0 & if \quad MS_W > MS_B \end{cases},$$

where $Count = n$ is the number of measurements in the group; MS_B – variance between groups;

- intermediate precision variance – $s_R^2 = s_r^2 + s_B^2$.

The results of checks, for example, a value of intermediate precision of the factors "time",

	A	B	C	D	E	F	G	H
80		Anova: Single Factor						
81								
82		SUMMARY						
83		*Groups*	*Count*	*Sum*	*Average*	*Variance*		
84		O1-D1	10	0,989559839	0,09895598	3,3E-07		
85		O1-D2	10	0,989191579	0,09891916	1,75E-07		
86		O2-D1	10	0,991524834	0,09915248	1,24E-07		
87		O2-D2	10	0,986950527	0,09869505	1,83E-07		
88								
89								
90		ANOVA						
91		*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*Fcrit*
92		Between Groups	1,0549E-06	3	3,5163E-07	1,732021	0,1777844	2,866265557
93		Within Groups	7,3087E-06	36	2,0302E-07			
94								
95		Total	8,3636E-06	39				

Fig. 5. ANOVA results

"operator", among other things, makes it possible in conjunction with a model approach to the assessment of the accuracy to obtain a more realistic evaluation of the measurement uncertainty in the laboratory [4].

3. Conclusion

1. Intralaboratory check can be carried out taking into account the need to identify the drift of the data obtained and using the analysis of variance.

2. If the volume of experimental data is not less than 100–300, the Poincaré plot can be used to analyze the quality of measurements.

3. The check results can be used to clarify the measurement uncertainty obtained with the model approach.

Забезпечення достовірності результатів калібрувань та випробувань статистичними методами

О.І. Колбасін [1,2]

[1] Державне підприємство "Харківський регіональний науково-виробничий центр стандартизації, метрології та сертифікації", вул. Мироносицька, 36, 61002, Харків, Україна
[2] Національний науковий центр "Інститут метрології", вул. Мироносицька, 42, 61002, Харків, Україна
kai_2548@i.ua

Анотація

Згідно з вимогами ISO/IEC 17025:2017, достовірність результатів випробувань та калібрувань забезпечується, серед іншого, валідацією застосовуваних методів та внутрішньолабораторною перевіркою отримуваних результатів. При цьому застосування статистичних методів має переваги при виявленні тенденцій та оцінюванні значущості можливих відхилень при варіюванні факторів, які впливають на результат вимірювання, але не піддаються повному контролю.

Стандарти ISO 5725 визначають низку таких методів, проте вибір конкретних методів залишається за лабораторією з огляду на вимоги щодо адекватності витрачених зусиль, ресурсів та часу цілям виконуваних робіт та ризикам отримання невідповідних результатів. При цьому лабораторія сама повинна певним чином визначити, які об'єкти калібрування (випробування) слід переважно використовувати в перевірках та яку періодичність перевірок слід передбачити.

У зв'язку з підвищенням точності та складності вимірювальних систем стає все більш очевидною необхідність застосування методів теорії випадкових процесів.

Показано, що використання діаграми Пуанкаре дозволяє комплексно, ефективно та наглядно оцінювати зміни у вимірювальному процесі з точки зору динаміки отримуваних результатів вимірювань.

Результати перевірок, наприклад, визначена проміжна прецизійність за факторами "час", "оператор", крім усього іншого, дає можливість у поєднанні з модельним підходом до оцінювання точності отримати більш реалістичну оцінку невизначеності вимірювань у лабораторії відповідно до ISO 21748.

У роботі аналізуються деякі практичні підходи (різного ступеня складності) до внутрішньолабораторних перевірок достовірності результатів калібрувань (випробувань), зокрема, з застосуванням ANOVA, Microsoft Office Excel та інших пакетів статистичних програм.

Ключові слова: калібрування; випробування; внутрішньолабораторна перевірка; діаграма Пуанкаре; ANOVA.

Обеспечение достоверности результатов калибровок и испытаний статистическими методами

А.И. Колбасин [1,2]

[1] Государственное предприятие "Харьковский региональный научно-производственный центр стандартизации, метрологии и сертификации", ул. Мироносицкая, 36, 61002, Харьков, Украина
[2] Национальный научный центр "Институт метрологии", ул. Мироносицкая, 42, 61002, Харьков, Украина
kai_2548@i.ua

Аннотация

Согласно требованиям ISO/IEC 17025:2017 достоверность результатов испытаний и калибровок обеспечивается среди прочего внутрилабораторной проверкой получаемых результатов.

Стандарты ISO 5725 определяют ряд статистических методов, однако выбор конкретных методов остается за лабораторией с учетом требований по адекватности затраченных усилий, ресурсов и времени целям выполняемых работ и рисков получения несоответствующих результатов.

В связи с повышением точности и сложности измерительных систем становится все более очевидной необходимость применения методов теории случайных процессов.

Показано, что использование диаграммы Пуанкаре позволяет комплексно и эффективно оценивать изменения в измерительном процессе с точки зрения динамики получения результатов измерений.

Промежуточная прецизионность, определенная по результатам проверок, дает возможность получить более реалистичную оценку неопределенности измерений в лаборатории в соответствии с ISO 21748.

В работе также анализируются некоторые практические подходы к внутрилабораторным проверкам достоверности результатов.

Ключевые слова: калибровка; испытания; внутрилабораторная проверка; диаграмма Пуанкаре; ANOVA.

References

1. ISO/IEC 17025:2017. General requirements for the competence of testing and calibration laboratories. 30 p.
2. Tested & Accepted Implementing ISO/IEC 17025:2017. United Nations Industrial Development Organisation (UNIDO), 2020. 72 p.
3. ISO 5725. Accuracy (trueness and precision) of measurement methods and results. Part 1 to Part 6.
4. ISO 21748:2017. Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation. 38 p.
5. JCGM 100:201X CD Committee Draft. Evaluation of measurement data – Guide to uncertainty in measurement.
6. Elyasberg P.E. Izmeritelnaya informaciya: skolko eyo nuzhno? Kak eyo obrabatyvat? [Measuring information: how much is it needed? How to process it?] Moscow, Nauka Publ., 1983. 208 p. (in Russian).
7. Adler Yu., Shper V. Statistical Thinking and Metrology: Problems and Decisions. *Metrology for Green Growth: proceedings of XX IMEKO World Congress*, September 9–14, 2012, Busan, Republic of Korea.
8. Charki A., Pavese F. Data comparisons and uncertainty: a roadmap for gaining in competence and improving the reliability of results. *International Journal of Metrology and Quality Engineering*, 2019, vol. 10, no. 1, pp. 1–10. doi: https://doi.org/10.1051/ijmqe/2018016
9. ISO 16269-4:2010(en). Statistical interpretation of data – Part 4: Detection and treatment of outliers.
10. ISO 5479:1997(en). Statistical interpretation of data – Tests for departure from the normal distribution.
11. Rutherford A. ANOVA and ANCOVA. A GLM Approach. John Wiley & Sons, 2011. 360 p.