ISBN: 978-628-95207-8-1. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2024.1.1.2040

Evaluation of the Calibration of Outside Micrometers by Statistical Repeatability Analysis

Ismael Correa-Gómez[1], Sergio Martínez-Campo[2], Libis Valdez-Cervantes[3], Rosa Martínez-Ospina[4], and, Luis Germán-Florez[5]
[1,2,4]Universidad Cooperativa de Colombia UCC, Colombia, Ismael.correag@campusucc.edu.co, sergio.martinezc@campusucc.edu.co, rosa.martinezo@campusucc.edu.co
[3]Fundación Universitaria Antonio de Arévalo UNITECNAR, Colombia, decano.fadi@unitecnar.edu.co
[5]Universidad del Atlántico, Colombia, lgerman@mail.uniatlantico.edu.co

Abstract– The study presented in the article focuses on the application of an intermediate accuracy test in a calibration laboratory to assess repeatability among authorized technical personnel performing calibration services in the length range. Technology watch was used to identify relevant, emerging and declining trends in repeatability assessment. To carry out the assessment, the procedure DI-005 established by the Spanish Metrology Centre (CEM) for the calibration of two-contact outside micrometers was followed. Hypothesis tests were applied to estimate the variance ratio and the differences of the population means, ensuring homogeneity of variance, normality (using the Shapiro-Wilk test) and the sample size necessary to use the t-student test. The results obtained in the research showed that the p-values were greater than the 5% significance level, and that the confidence interval estimation for the limits included zero, indicating equality of population means. This demonstrated the existence of repeatability among authorized technical personnel, confirming the reliability of the measurements and the ability to issue calibration certificates.

Keywords-- Repeatability, calibration, technology watch, inferential statistics, homogeneity of variance, normality.

I. INTRODUCTION

In our industrialized society, many aspects of our daily lives depend on precise measurements. With the advancement of modern techniques, there is an increasing demand for accuracy, range and diversity of standards in different areas. Metrology plays a crucial role in the marketing of products nationally and internationally, as quality assurance and customer confidence agencies use metrology to ensure that products meet minimum quality standards anywhere in the world [1]. Knowledge about the application of metrology is essential in all scientific professions, as it enables the quantitative measurement of the physical and chemical properties of objects. Metrology deals with the study and application of different methods for measuring quantities such as lengths, angles, masses, times, velocities, powers, currents, temperatures and pressures. Throughout history, scientific progress has been closely linked to advances in measurement capability [1].

Ensuring an adequate level of measurement accuracy is a way to prevent incorrect decisions and reduce economic losses associated with defective products and processes. It also improves the management and operation of measurement activities in organisations, providing greater confidence in claims of conformity with product and service requirements, as well as regulatory requirements [2].

An effective measurement management system must ensure that measurement equipment and processes are fit for purpose, that product quality objectives are achieved and that the risk of incorrect measurement results is managed [2]. To improve quality, a reliable Measurement System is essential. For measurements and tests, it is crucial to have measurement standards that are reproducible. It is important to bear in mind that measurement and test results are always subject to some uncertainty [3]. Therefore, the calibration operation under specific conditions consists of two stages, the first stage is a relationship between the values and their measurement uncertainties obtained from the measurement standards and the corresponding indications with their associated uncertainties, and in a second stage, it uses this information to establish a relationship that allows obtaining a measurement result from an indication [1].

However, the evaluation of measurement systems used to focus on characteristics such as accuracy, linearity and stability of equipment, instruments or devices. It is now recognised that it is important to also include repeatability and reproducibility in these evaluations, as these are fundamental properties of measurement systems [3]. The study of repeatability and reproducibility is applied in the evaluation of proficiency tests, validation of calibration methods, analysis of interlaboratory comparisons, evaluation of measurement uncertainty, evaluation of control charts, variability of measurements and instruments, evaluation of instrument drift (stability), among others [4].

Accepted methods for the determination of repeatability and reproducibility studies are based on the statistical evaluation of the dispersions of the results, either in the form of statistical range (maximum-minimum) or their representation as variances or standard deviations, these methods are: Mean Range and ANOVA (analysis of variance) [4].

This paper presents a research study that seeks to determine the variability of measurements between two technicians by applying an intermediate precision test to determine the repeatability between them, based on results obtained through hypothesis testing. This intermediate precision test will allow the laboratory to evaluate the performance of the technical staff to know how repeatable the measurements are between them, guaranteeing the validity of the results when issuing calibration certificates. The reliability of this method is associated with compliance with procedure

DI-005 for the calibration of two-contact outside micrometers, established by the CEM, whose purpose is to guarantee reliable results to users, since if they are not repeatable the method cannot be used.

II. Conceptual framework

Measurements play an important role in people's daily lives. They are found in every activity, from the naked eye estimation of a distance to a control process or basic research [1]. Metrology laboratories must meet certain installation requirements, so that at any time data can be determined with the greatest possible confidence, measurements, and verifications with the smallest possible error [2]. The performance of intermediate precision (reproducibility) and repeatability tests within the laboratory is an important step in the process of uncertainty estimation or method validation. These tests should be included to comply with NTC-ISO/IEC 17025:2017 [2].

The Colombian Institute of Technical Standards and Certification (ICONTEC) in the standard NTC-ISO/IEC 17025:2017, determines that the laboratory must have a procedure for monitoring the validity of the results. The resulting data should be recorded in such a way that trends are detectable and where possible, statistical techniques should be applied for the review of results, as is the case for an intermediate precision test among laboratory technicians [2].

However, to assess the degree of repeatability of a given measurement, the researcher must measure the same trait, on the same individual, at least twice. The second measurement should obviously be made without remembering or checking what the value obtained in the first measurement was. A good method may be to measure 10-20 individuals the first time, mark them individually, place them in a cage or bag, and measure them again by removing them at random, noting the new measurement values on a separate sheet of paper [3].

Consequently, two different technicians may report the result of a test differently. Therefore, it is important to achieve a consensus between observers, and a useful tool to unify the observation criteria is the transformation of the observers' assessments into quantitative terms [4]. Repeatability is also related to precision. It refers to measurements obtained under nearly identical conditions [5].

A. Technology Surveillance

Technology Surveillance is one of the main tools of the Technology Intelligence System, such as technological diagnosis, performance analysis, organizational and institutional benchmarking, commercial surveillance and technological foresight, which have the objective of generating useful and strategic knowledge through the search, management and analysis of information [6]. Currently, the scientific-technological paradigm is consolidating, in which the generation of information and the way in which it is treated is of great importance, so the central postulate is its strategic management [6]. Thus, technology watch is defined as a systematic process in which useful information is captured, analysed and exploited to contribute to strategic decision-making, consistent with the environment [9]. Below, Fig. 1 details a scheme that shows the existing relationships between traditional metrics, the new techniques for measuring information and surveillance, as well as the resources that serve as input for the process [7].

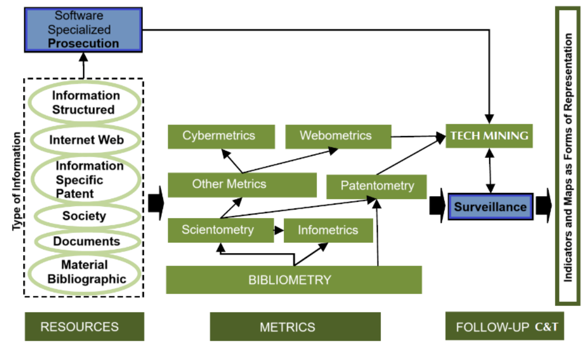

Fig. 1 Overview of metrics, technologies and surveillance.

Bibliometrics is the root of all sciences that rely on the treatment of information to establish dynamics in each of them; therefore, bibliometric indicators are used to describe the behavior of academic communities, as is the case of scientometric indicators, within which, by means of technological maps, information related to books or articles is analyzed, and also patents as a strategic part of innovation, giving way to the systematization and technification of processes [7].

Fig. 2 describes the monitoring procedure in four stages, as follows: I. Planning and identification of needs; II. Identification, search and collection of information; III. Organising, cleaning and analysing the information; and IV. Communication, decision-making and use of the results obtained in the process [7].

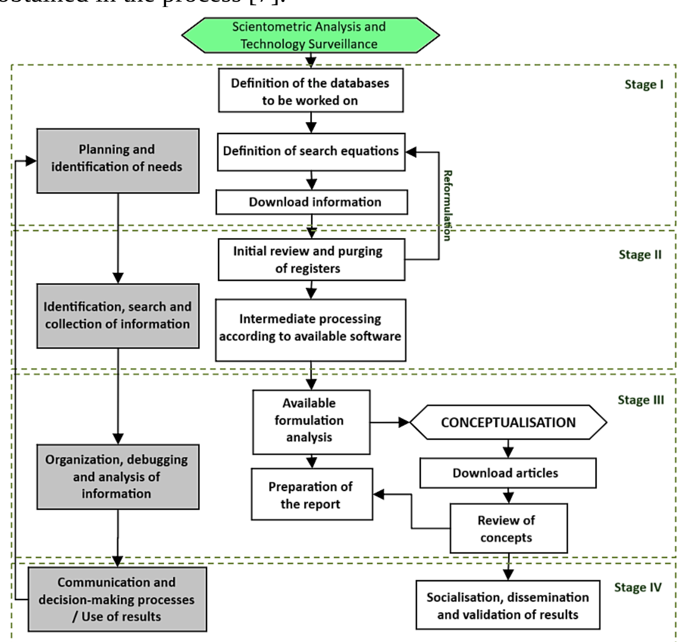

Fig. 2 Technology Surveillance process.

Applying the above procedure, the following databases to be used are defined: Scopus, Science Direct, OpenAlex, Lens and Dimensions, and the search formula or equation and the results obtained are generated, as shown in Table I.

TABLE I
BIBLIOGRAPHIC MAPPING

Database	**Documents Articles**	**Equation**	**Timespan**
Scopus	48	(TITLE-ABS-KEY (repeatability AND studies AND ion AND measurement AND laboratories) AND TITLE (repeatability AND measurement))	2015:2024
Science Direct	89	Repeatability studies in measurement laboratories Title: Repeatability and measurement	2010:2024
OpenAlex	100	Search title, abstract, & test: “Repeatability”. Type: article. Concept: Calibration & Accuracy and precision & metrology	2000:2024
Lens	117	Obras académicas (117) = Title: (micrometer AND external) OR (Abstract: (Repeatability AND (measurement AND micrometer)) OR (Keyword: (Repeatability AND (measurement AND micrometer)) OR (Field of Study: (Repeatability AND (measurement AND micrometer)) OR (Title: measurement OR (Title: Repeatability OR Field of Study: measurement)))))	2012:2024
Dimensions	775	Publication Type: Article, Free text in title and abstract: "Repeatability and reproducibility" and "measurement"	2017:2024

Source: Own elaboration

The bibliometric analysis is carried out through the study of the results produced by the Biblioshiny/Bibliometrix interface, as a statistical tool for quantitative research based on scientometric and bibliometric studies.

From the publication records analysed in the Technology Surveillance using the specialised software, the top 5 of the most relevant articles were selected for each database, where the formation of five knowledge clusters is evident. The publication dynamics of each cluster are presented in Table II.

As a result of the above analysis, technology watch assists decision making for the implementation of repeatability as an assessment of measurement uncertainty in a laboratory, which must take into account all contributions that are significant, including those arising from sampling, calibration, using appropriate methods of analysis.

TABLE II
PROFILE OF REPEATABILITY KNOWLEDGE CLUSTERS

Clúster	**Artículos**	**DataBase**				
		SCOPUS	Science Direct	OpenAlex	Lens	Dimensions
(1) Measurement and instrument variability – Repeatability.	[8], [9]	X				
	[27]				X	
(2) Variability of measurements and instruments - Repeatability and Reproducibility	[10], [11]	X				
(3) Assessment of measurement uncertainty - Repeatability	[12]	X				
	[13], [14], [15]		X			
	[18], [19], [20], [21], [22]			X		
	[23], [24], [25], [26]				X	
	[28], [29]					X
(4) Measurement Uncertainty Assessment - Repeatability and Reproducibility	[16], [17]		X			
	[30]					X
(5) Validation of Calibration Methods - Repeatability and Reproducibility	[31], [32]					X

Source: Own elaboration

B. Repeatability and reproducibility concepts

Repeatability and reproducibility studies of measurements determine that part of the variation observed in the process is due to the measurement system or method used [3]. Repeatability can be expressed quantitatively in terms of the characteristic dispersion of the results. Repeatability is defined, according to the VIM (International Vocabulary of Metrology), as the closeness of agreement between the results of successive measurements of the same measurand under the same measurement conditions, including: same measurement procedure, same technician/observer, same measuring instrument, used under the same conditions, same location, repeated over a short period of time [4].

Whereas reproducibility is defined as the closeness of agreement between the results of successive measurements of the same measurand under changing measurement conditions. A valid reproducibility statement requires that the changing condition be specified. Changing conditions may include: measurement principle, measurement method, technician/observer, measuring instrument, reference standard, location, conditions of use, time. Reproducibility can be expressed quantitatively in terms of the characteristic dispersion of results [4].

C. Theory of Errors

In carrying out a measurement process it is not possible to avoid a number of errors, but it is possible to keep them to a minimum. Measurement errors by the technician are of course unavoidable, but they can be reduced by practice, so that the technician in his measuring function should take care to make them as little as possible. The main errors that can be made are as follows [1]:

Parallelism error. This results from the incorrect position of the technician to read the reading indicated by the device, the recommended way is for the technician to position himself perpendicular to the scale or dial where the reading should be taken. Fig. 3 shows the reading positions, where (b) is the correct one [1].

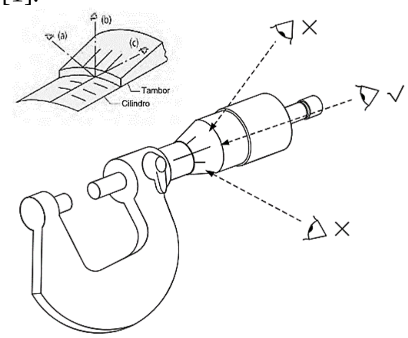

Fig. 3 Parallelism error

Pressure error. This occurs when the apparatus or instrument lacks in its construction, some element that neutralizes or regulates an excess of effort used in the handling of the apparatus. In the measurement itself, it should not be forgotten that, if the action is carried out with greater or lesser effort, a measurement reading of a different value will be produced, which will depend on the degree of effort used due to flattening of the contact surfaces of the instrument. It is also the case, when gauges are used, that when they are held manually with greater force than necessary, the sensitivity decreases [1].

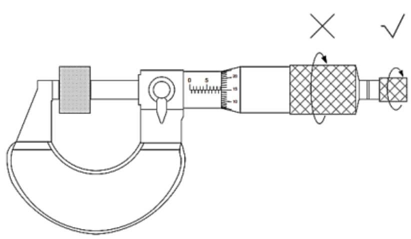

Fig. 4 Pressure error when tightening and the correct way to do it

Fig. 4 describes how to use the ratchet to squeeze the object to be measured between the measuring stops. It is recommended to make the sound produced by the ratchet three times when tightening.

Micrometer on Stand Positioning Error. When mounting the micrometer on a stand, care should be taken to ensure that the micrometer body is supported by the center of the micrometer frame and that the support is not too strong, as shown in Fig. 5.

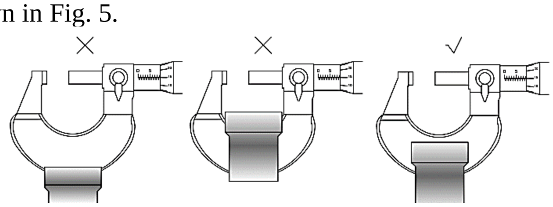

Fig. 5 Placing the micrometer on a stand

Origin or Zero Error. When the micrometer is closed and the reading does not indicate zero. It is important to check that the zero indications coincide when the measuring stops are in contact.

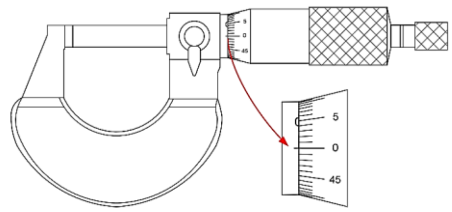

Fig. 5 Origin or Zero Error

III. Methodology

A. Intermediate precision test execution.

Intermediate precision within the laboratory is important as it demonstrates that measurement results can be reproduced under different conditions, i.e. the results will be the same whether the tests or calibrations are performed by analyst 1 or analyst 2.

The most common test to assess intermediate precision involves two or more analysts making repeated measurements.

To perform the test, follow these steps for each analyst:

1. Perform a repeatability test with the analyst.
2. Record the results.
3. Calculate the measurement, standard deviation, and degrees of freedom from the results obtained in the previous step.
4. Calculate the standard deviation of the mean values obtained in the previous step, for each analyst.

A description is given of each of the stages involved in the execution of the intermediate precision test [33].

Selection of personnel. The personnel selected for this test must have prior knowledge of the handling and use of the measuring instrument to be used. Selection of the measuring instrument and part to be measured. The instrument and the part must be selected in advance, to be subjected to a readiness process.

Enlistment process. The instrument and the part are subjected to a cleaning process using isopropyl alcohol and rice paper, eliminating any type of dirt that may alter the results obtained in the test, and then undergo a tempering process.

Tempering process. The instrument and the part are left to temper for 24 hours before the test is carried out, so that both the instrument and the part reach thermal equilibrium at a temperature of 20 ± 2°C.

Assembly process. The outside micrometer is placed on a specialized base for this measuring instrument, facilitating its handling, before taking the data.

Data acquisition process. The part is placed between the measuring faces of the two-contact outside micrometer and data is taken by each of the laboratorians taking part in this test, with a total of 20 data obtained by each one.

Data processing. The data obtained in the measurement of the samples are subjected to verification of assumptions such as homogeneity of variances (Fisher) and normality (Shapiro-Wilk). Subsequently, outliers are verified (box-and-whisker plot), application of the t-student test and estimation by confidence intervals, to demonstrate the difference of equality of means.

B. *Materials and methods*

This study was carried out in a calibration metrology laboratory in the city of Barranquilla, Atlántico - Colombia, carrying out a series of measurements between two technicians, who took the data of the measurement obtained with the same instrument, in this case a two-contact outside micrometer, and the same piece under controlled conditions, in order to corroborate the similarity in the results, which are expressed in millimeters (mm). Based on the above, we compare the present study to experimental-type research.

In this research, non-probabilistic sampling was carried out with a sample size of four (4), in 5 series (runs), where four (4) variables were analyzed, two qualitative (Position and Series) and two quantitative (Team Measurement and Experience).

Below, in Table III, the variables to be measured in order to carry out the test are presented:

TABLE III
METADATA TABLE.

Metadata table			
Variable	**Label**	**Type**	**Unit of measurement**
Position	Description	Qualitative	one and two
Equipment measurement	Two-contact outside micrometer	Quantitative	mm
Series	Run	Qualitative	Integers
Experience	Years of experience	Quantitative	Years

Source: Own elaboration

IV. RESULTS

Using a sample size four (4) in five (5) runs with time intervals of five minutes, the two metrology laboratory technicians proceeded to take the measurements of the outside micrometer, the data of which are as follows:

TABLE IV
TECHNICIAN 1 DATA.

Run	1	2	3	4
1	10,300	10,303	10,301	10,300
2	10,300	10,301	10,300	10,299
3	10,298	10,300	10,298	10,302
4	10,300	10,299	10,301	10,301
5	10,302	10,303	10,300	10,302

Source: Own elaboration

TABLE V
TECHNICIAN 2 DATA.

Run	1	2	3	4
1	10,302	10,302	10,300	10,298
2	10,301	10,298	10,301	10,299
3	10,300	10,298	10,300	10,302
4	10,302	10,299	10,301	10,300
5	10,301	10,301	10,303	10,299

Source: Own elaboration

Table VI shows a descriptive analysis [34] where the average, standard deviation, and coefficient of variation for each sample may be observed, with the support of the InfoStat Statistical Software [35], which results are shown below:

TABLE VI
DESCRIPTIVE ANALYSIS

Variable	**n**	**Average**	**Var(n-1)**	**CV**
C1T-two	4	10,301	3.7E-06	0.019
C2T-two	4	10,300	2.3E-06	0.015
C3T-two	4	10,300	2.7E-06	0.016
C4T-two	4	10,301	1.7E-06	0.013
C5T-two	4	10,301	2.7E-06	0.016
C1T-one	4	10,301	2.0E-06	0.014
C2T-one	4	10,300	6.7E-07	0.008
C3T-one	4	10,300	3.7E-06	0.019
C4T-one	4	10,300	9.2E-07	0.009
C5T-one	4	10,302	1.6E-06	0.012

Source: Own elaboration

A. *Test for homogeneity of variance*

From the measures calculated in Table VI, we can test the homoscedasticity of the variances using the two-variance ratio test, where S_1^2 and S_2^2 are the variances of the independent samples of size n_1 y n_2 taken from normal populations σ_1^2 and σ_2^2, respectively [36].

The hypotheses will be tested, which will compare the homogeneity of variance of one population with the other.

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Applying the Fisher test statistic for each of the five (5) runs taking into account the years of experience of the technicians. The data of technician two are placed in the numerator because he has more years of experience, wherein the test statistic has the form:

$$F = \frac{s_2^2}{s_1^2}$$

and corresponds to the value of a random variable that has an F distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ with degrees of freedom. Table VII shows the obtained results.

TABLE VII
FISHER TEST RESULTS

No. of Runs	**Statistical**
1	F = 1.833
2	F = 3.375
3	F = 0.727
4	F = 1.818
5	F = 1.684

Source: Own elaboration

When comparing the value of the theoretical Fisher which is 9.27 coming from the significance level of 0.05 and degrees of freedom 3;3, we observe that the Fisher calculated for each of the five runs are lower, therefore, it is inferred that there is no significant evidence to reject the null hypothesis, that is, the variances of the samples from each run are equal.

B. *Atypical data test*

We will check if there are atypical data in each of the samples. Initially we will graph the box and whisker plots for each of the samples of the ten runs using the InfoStat Statistical Software [35].

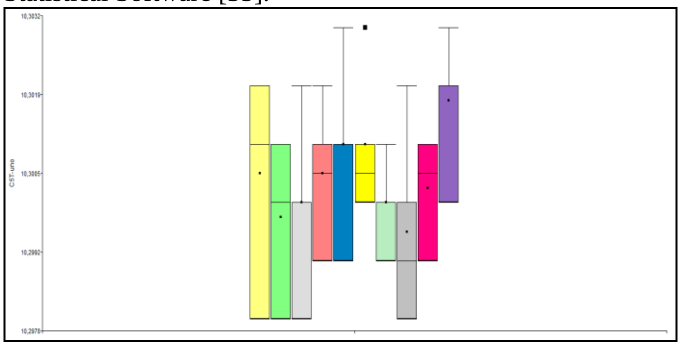

Fig. 6 Box and whisker plot by runs.

It is observed in the box and whisker plot diagram number 6 (yellow color), corresponding to run one of technician one, a data above the upper whisker plot, which leads us to determine that there are atypical data.

To do this, the first quartiles (Q1) and the third quartiles (Q3) are initially calculated for each of the samples of the ten runs using the InfoStat statistical software and the interquartile range (IQR), through the difference between Q3 and Q1, in Microsoft Excel.

TABLE VIII
VALUE OF THE Q1 AND Q3 QUARTILES

Variable	n	Average	Q1	Q3
C1T-two	4	10,301	10,298	10,302
C2T-two	4	10,300	10,298	10,301
C3T-two	4	10,300	10,298	10,300
C4T-two	4	10,301	10,299	10,301
C5T-two	4	10,301	10,299	10,301
C1T-one	4	10,301	10,300	10,301
C2T-one	4	10,300	10,299	10,300
C3T-one	4	10,300	10,298	10,300
C4T-one	4	10,300	10,299	10,301
C5T-one	4	10,302	10,300	10,302

Source: Own elaboration

TABLE IV
CALCULATION OF THE INTERQUARTILE RANGE

RIC C1T-two	RIC C2T-two	RIC C3T-two	RIC C4T-two	RIC C5T-two
0.004	0.003	0.002	0.002	0.002
RIC C1T-one	**RIC C2T-one**	**RIC C3T-one**	**RIC C4T-one**	**RIC C5T-one**
0.001	0.001	0.002	0.002	0.002

Source: Own elaboration

In the box and whisker plot, observations with scores greater than 1.5*IQR evaluated from Q75 (Q3) or less than 1.5*IQR evaluated from Q25 (Q1) are considered atypical data. From 3*RIC they are classified as extreme [37].

TABLE X
ATYPICAL DATA LIMITS

Atypical C1T-two	Atypical C2T-two	Atypical C3T-two	Atypical C4T-two	Atypical C5T-two
10.2920	10.2935	10.2950	10.2960	10.2960
10.3080	10.3055	10.3030	10.3040	10.3040
Atypical C1T-one	**Atypical C2T-one**	**Atypical C3T-one**	**Atypical C4T-one**	**Atypical C5T-one**
10.2985	10.2975	10,295	10,296	10,297
10.3025	10.3015	10,303	10,304	10,305

Source: Own elaboration

When comparing the data of the samples from each of the ten runs, it is observed that data 10.303 from run one of technician one (C1T-one) is outside the limits (10.2985: 10.3025), which is determined as an atypical data.

C. *Data Normality Test*

To test the normality of the data, the Shapiro-Wilk test was used, because the sample size is four (4) less than fifty. We will apply this test for each of the ten runs using the statistical software, at a significance level of 5%.

D. *The following hypotheses are proposed:*

$$H_0: \quad X \sim N(\mu, \sigma^2)$$
$$H_1: \quad X \nsim N(\mu, \sigma^2)$$

TABLE XI
SHAPIRO-WILK STATISTIC.

Variable	n	Average	OF	W *	P-value
C1T-two	4	10,301	1.9E-03	0.85	0.2462
C2T-two	4	10,300	1.5E-03	0.83	0.1795
C3T-two	4	10,300	1.6E-03	0.96	0.7298
C4T-two	4	10,301	1.3E-03	0.99	0.9293
C5T-two	4	10,301	1.6E-03	0.96	0.7298
C1T-one	4	10,301	1.4E-03	0.82	0.1574
C2T-one	4	10,300	8.2E-04	0.96	0.7298
C3T-one	4	10,300	1.9E-03	0.85	0.2462
C4T-one	4	10,300	9.6E-04	0.85	0.2462
C5T-one	4	10,302	1.3E-03	0.91	0.4808

Source: Own elaboration

In Table VI it is observed that the values in the column corresponding to the p-value of the ten runs are all greater than the significance level of 0.05, so there is no significant evidence to reject the null hypothesis, that is, the data is distributed normally.

E. *Test for difference of two averages and confidence intervals*

To estimate by intervals the differences in population average of technician one with respect to technician two, the t-student test was applied because the population variances are unknown and equal with small (size 4) and independent samples and the distribution of the data is normal with a confidence level of 95% [38].

The following hypotheses are established:

μ_1: It is the population average of all technicians one runs

μ_2: It is the population average of all technicians two runs

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

We compared the measurements as follows:

Run one of technician one versus run one of technician two.

TABLE XII

SHAPIRO-WILK STATI STUDENT T-TEST STATISTICS FOR RUN 1

Classify	Variable	Group 1	Group 2	n(1)	n(2)	Average (1)	
Technical	C1	1	2	4	4	10.3010	
Average (2)	**M1-M2**	**LI(95)**	**LS(95)**	**T**	**gl**	**p-value**	**Proof**
0.3005	0.0005	-0.0024	0.0034	0.4201	6	0.6891	Bilateral

Source: Own elaboration

Run two of technician one versus run two of technician two.

TABLE XIII

SHAPIRO-WILK STATI STUDENT T-TEST STATISTICS FOR RUN 2

Classify	Variable	Group 1		Group 2	n(1)	n(2)	Average (1)
Technical	C1	1		2	4	4	10,300
Average (2)	**M1-M2**	**LI(95)**	**LS(95)**	**T**	**gl**	**p-value**	**Proof**
10.2998	0.0002	-0.0018	0.0023	0.2928	6	0.7796	Bilateral

Source: Own elaboration

Run three of technician one versus run three of technician two.

TABLE XIV

SHAPIRO-WILK STATI STUDENT T-TEST STATISTICS FOR RUN 3

Classify	Variable	Group 1	Group 2		n(1)	n(2)	Average (1)
Technical	C1	1	2		4	4	10.2995
Average (2)	**M1-M2**	**LI(95)**	**LS(95)**	**T**	**gl**	**p-value**	**Proof**
10,300	-,0005	-0.0036	0.0026	-0.3974	6	0.7049	Bilateral

Source: Own elaboration

Run four of technician one versus run four of technician two.

TABLE XV

SHAPIRO-WILK STATI STUDENT T-TEST STATISTICS FOR RUN 4

Classify	Variable	Group 1	Group 2	n(1)	n(2)	Average (1)	
Technical	C1	1	2	4	4	10.3003	
Average (2)	**M1-M2**	**LI(95)**	**LS(95)**	**T**	**gl**	**p-value**	**Proof**
10.3005	-,0002	-0.0022	0.0017	-0.3111	6	0.7663	Bilateral

Source: Own elaboration

Run five of technician one versus run five of technician two.

TABLE XVI

SHAPIRO-WILK STATI STUDENT T-TEST STATISTICS FOR RUN 5

Classify	Variable	Group 1	Group 2		n(1)	n(2)	Average (1)
Technical	C1	1	2		4	4	10.3018
Average (2)	**M1-M2**	**LI(95)**	**LS(95)**	**T**	**gl**	**p-value**	**Proof**
10.3010	.0007	-0.0018	0.0033	0.7276	6	0.4943	Bilateral

Source: Own elaboration

In the previous tables, it is shown that all the p-values are greater than the 0.05 level of significance, which proves that there is no significant evidence to reject the null hypothesis that the averages are equal. Likewise, the estimation of the population average was carried out using the confidence intervals, where it is observed that the lower limits (LI) and upper limits (LS) include zero (0), which also proves that the averages are equal, concluding that the average of all runs are equal.

V. CONCLUSIONS

According to the study described above, it was proven that there is repeatability among the technicians of the calibration laboratory under study, for the provision of the calibration service of two-contact outside micrometers, ensuring the validity of the results at the time of issuing certificates because the coefficients of variation (CV) tend to zero (0), which indicates low variability between the measurements of each technician.

By applying the variance ratio test, it was proven that there is repeatability between the technicians, having considerable precision between them, since the calculated Fisher were lower than the theoretical Fisher. In the atypical data test, it was observed that in run one of technician one, there is an atypical data (10,303 mm), which did not influence the conclusion of the study. Using the Shapiro-Wilk test, it was confirmed that the data follow a normal distribution because the p-value of each of the runs was greater than the 5% significance level. By using the t- student test and estimating the means by intervals, it was verified that the means of the technicians' measurements are equal, the calculated p-values were greater than 5% and the limits of the confidence interval include zero.

REFERENCES

[1] DI-005 Procedimiento para la calibración de micrómetros de exteriores de dos contactos | Centro Español de Metrología. (s. f.). Centro Español de Metrología. https://www.cem.es/es/divulgacion/documentos/di-005-procedimiento-calibracion-micrometros-exteriores-dos-contactos.html.

[2] Requisitos generales para la competencia de los laboratorios de ensayo y calibración. (s. f.). https://tienda.icontec.org/gp-requisitos-generales-para-la-competencia-de-los-laboratorios-de-ensayo-y-calibracion-ntc-iso-iec17025-2017.html.

[3] Senar, J. C. (1999). La medición de la repetibilidad y el error de medida. ResearchGate. https://www.researchgate.net/publication/233801237_La_medicion_de_la_repetibilidad_y_el_error_de_medida

[4] Donis, J. H., (2012). Evaluación de la validez y confiabilidad de una prueba diagnóstica. Avances en Biomedicina, 1(2), 73-81.

[5] Reinaldo Fretes, Vicente, Pedrozo, Andrea, Gamarra, José, Escobar, Patricia María, Cubilla, Raúl Enrique, & Adorno, Carlos Gabriel. (2019). Estudio preliminar sobre la repetibilidad in vivo de tres localizadores apicales electrónicos. Revista Cubana de Estomatología, 56(3), e2176. Epub 15 de octubre de 2019. Recuperado en 05 de mayo de 2024, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0034-75072019000300011&lng=es&tlng=es.

[6] Campo, S. M., Rada, S., & Nieto, L. (2023). Technological Surveillance of the Mathematical Modeling of COVID-19's Dynamics and Containment. In 2023 World Engineering Education Forum-Global Engineering Deans Council (WEEF-GEDC) (pp. 1-9). IEEE.

[7] González, D., Campo, S. M., Valdez, L., & Pinzón, M. (2023). Technological Surveillance on Social Insertion Strategies Aimed at the Female Prison Population.

[8] Washabaugh, E. P., Kalyanaraman, T., Adamczyk, P. G., Claflin, E. S., & Krishnan, C. (2017). Validity and repeatability of inertial measurement

units for measuring gait parameters. Gait & Posture, 55, 87-93. https://doi.org/10.1016/j.gaitpost.2017.04.013.
[9] Moerenhout, B. A., Gelaude, F., Swennen, G., Casselman, J., Van Der Sloten, J., & Mommaerts, M. Y. (2009). Accuracy and repeatability of cone-beam computed tomography (CBCT) measurements used in the determination of facial indices in the laboratory setup. Journal Of Cranio-maxillofacial Surgery/Journal Of Cranio-maxillo-facial Surgery, 37(1), 18-23. https://doi.org/10.1016/j.jcms.2008.07.006.
[10] Penna, D., Stenni, B., Šanda, M., Wrede, S., Bogaard, T., Gobbi, A., Borga, M., Fischer, B., Bonazza, M., & Chárová, Z. (2010). On the reproducibility and repeatability of laser absorption spectroscopy measurements for δ2H and δ18O isotopic analysis. Hydrology And Earth System Sciences, 14(8), 1551-1566. https://doi.org/10.5194/hess-14-1551-2010
[11] Balint PV, Sturrock RD. Intraobserver repeatability and interobserver reproducibility in musculoskeletal ultrasound imaging measurements. Clinical and Experimental Rheumatology. 2001 Jan-Feb;19(1):89-92. PMID: 11247333.
[12] Schall, M. C., Fethke, N. B., Chen, H., Oyama, S., & Douphrate, D. I. (2015). Accuracy and repeatability of an inertial measurement unit system for field-based occupational studies. Ergonomics, 59(4), 591-602. https://doi.org/10.1080/00140139.2015.1079335.
[13] Saadé, M., Mekhael, E., Nassim, N., Karam, M., Massaad, A., Jaber, E., Ayoub, E., Rteil, A., Rachkidi, R., & Assi, A. (2022). Repeatability of muscle strength measurements in patients with spinal deformity. Gait & Posture, 97, S148. https://doi.org/10.1016/j.gaitpost.2022.07.098.
[14] Grant, R., Coopman, K., Mayer, S., Kara, B., Campbell, J. A., Braybrook, J., & Petzing, J. N. (2018). Assessment of operator variation in flow cytometry measurements using gauge repeatability & reproducibility techniques. Cytotherapy, 20(5), S77. https://doi.org/10.1016/j.jcyt.2018.02.216.
[15] Lastra, P., Ripoll, C., Rincón, D., Catalina, V., & Bañnares, R. (2010b). 193 REPEAT MEASUREMENTS DO NOT ADD FURTHER INFORMATION TO SINGLE MEASUREMENTS IN THE PREDICTION OF OUTCOMES IN CIRRHOSIS. Journal Of Hepatology, 52, S83-S84. https://doi.org/10.1016/s0168-8278(10)60195-x.
[16] Hon, Y., Cheung, S., & Cho, P. (2011). 33 Repeatability of corneal biomechanical measurements in children wearing spectacles and orthokeratology lenses. Contact Lens & Anterior Eye/Contact Lens And Anterior Eye, 34, S24. https://doi.org/10.1016/s1367-0484(11)60112-4.
[17] Vermeulen, F., Vermeulen, F., Feyaerts, N., Proesmans, M., & De Boeck, K. (2009). Comparison of nasal potential difference measurements done on the nasal floor and under the inferior turbinate: repeatability. Journal Of Cystic Fibrosis, 8, S10. https://doi.org/10.1016/s1569-1993(09)60043-6.
[18] Edwards, H., McGlothlin, R., & Elisa, U. (1998). Vertical metrology using scanning-probe microscopes: Imaging distortions and measurement repeatability. Journal Of Applied Physics, 83(8), 3952-3971. https://doi.org/10.1063/1.367151.
[19] Bobroff, N. (1993). Recent advances in displacement measuring interferometry. Measurement Science & Technology, 4(9), 907-926. https://doi.org/10.1088/0957-0233/4/9/001.
[20] Kiekens, K., Welkenhuyzen, F., Tan, Y., Bleys, P., Voet, A., Kruth, J., & Dewulf, W. (2011). A test object with parallel grooves for calibration and accuracy assessment of industrial computed tomography (CT) metrology. Measurement Science & Technology, 22(11), 115502. https://doi.org/10.1088/0957-0233/22/11/115502.
[21] Wang, Z., & Maropolous, P. G. (2013). Real-time error compensation of a three-axis machine tool using a laser tracker. The International Journal Of Advanced Manufacturing Technology/International Journal, Advanced Manufacturing Technology, 69(1-4), 919-933. https://doi.org/10.1007/s00170-013-5019-5.
[22] Li, D., Tong, Z., Jiang, X., Blunt, L., & Gao, F. (2018). Calibration of an interferometric on-machine probing system on an ultra-precision turning machine. Measurement, 118, 96-104. https://doi.org/10.1016/j.measurement.2017.12.038.
[23] Sánchez, M., Castro, J. R., Castillo, O., Mendoza, O., Rodríguez-Díaz, A., & Melín, P. (2016). Fuzzy higher type information granules from an uncertainty measurement. Granular Computing, 2(2), 95-103. https://doi.org/10.1007/s41066-016-0030-5.
[24] Bakker LA, Schröder CD, van Es MA, Westers P, Visser-Meily JMA, van den Berg LH. Assessment of the factorial validity and reliability of the ALSFRS-R: a revision of its measurement model. J Neurol. 2017 Jul;264(7):1413-1420. doi: 10.1007/s00415-017-8538-4. Epub 2017 Jun 12. PMID: 28608303; PMCID: PMC5502060.
[25] Crabolu, M., Pani, D., Raffo, L., Conti, M., Crivelli, P., & Cereatti, A. (2017). In vivo estimation of the shoulder joint center of rotation using magneto-inertial sensors: MRI-based accuracy and repeatability assessment. BioMedical Engineering Online, 16(1). https://doi.org/10.1186/s12938-017-0324-0.
[26] Pavlovčič, U., Diaci, J., Možina, J., & Jezeršek, M. (2015). Wound perimeter, area, and volume measurement based on laser 3D and color acquisition. BioMedical Engineering Online, 14(1). https://doi.org/10.1186/s12938-015-0031-7.
[27] González, E., García, A. C., & Ventura, J. (2016). Weight Constrained DEA Measurement of the Quality of Life in Spanish Municipalities in 2011. Social Indicators Research, 136(3), 1157-1182. https://doi.org/10.1007/s11205-016-1426-y.
[28] Chiavaioli, F., Gouveia, C., Jorge, P. A. S., & Baldini, F. (2017). Towards a Uniform Metrological Assessment of Grating-Based Optical Fiber Sensors: From Refractometers to Biosensors. Biosensors, 7(4), 23. https://doi.org/10.3390/bios7020023.
[29] Yokoo, T., Serai, S. D., Pirasteh, A., Bashir, M. R., Hamilton, G., Hernando, D., Hu, H. H., Hetterich, H., Kühn, J., Kukuk, G., Loomba, R., Middleton, M. S., Obuchowski, N. A., Song, J. S., Tang, A., Wu, X., Reeder, S. B., & Sirlin, C. B. (2018). Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis. Radiology, 286(2), 486-498. https://doi.org/10.1148/radiol.2017170550.
[30] Shukla-Dave, A., Obuchowski, N. A., Chenevert, T. L., Jambawalikar, S., Schwartz, L. H., Malyarenko, D. I., Huang, W., Noworolski, S. M., Young, R. J., Shiroishi, M. S., Kim, H. M., Coolens, C., Laue, H., Chung, C., Rosen, M., Boss, M. A., & Jackson, E. F. (2018). Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. Journal Of Magnetic Resonance Imaging, 49(7). https://doi.org/10.1002/jmri.26518.
[31] Lei, J., Durbin, M. K., Shi, Y., Uji, A., Balasubramanian, S., Baghdasaryan, E., Al-Sheikh, M., & Sadda, S. R. (2017). Repeatability and Reproducibility of Superficial Macular Retinal Vessel Density Measurements Using Optical Coherence Tomography Angiography En Face Images. JAMA Ophthalmology, 135(10), 1092. https://doi.org/10.1001/jamaophthalmol.2017.3431.
[32] Popović, Z. B., & Thomas, J. D. (2017). Assessing observer variability: a user's guide. Cardiovascular Diagnosis And Therapy, 7(3), 317-324. https://doi.org/10.21037/cdt.2017.03.12
[33] Vocabulario Internacional de Metrología VIM, 3a edición 2012 (español) | Centro Español de Metrología. (s. f.). Centro Español de Metrología. https://www.cem.es/es/divulgacion/documentos/vocabulario-internacional-metrologia-vim-3a-edicion-2012-espanol.
[34] Posada, G. J., y Buitrago, M. V. (2016). Elementos básicos de estadística descriptiva para el análisis de datos [recurso electrónico]. Colombia: Fundación Universitaria Luis Amigó.
[35] Di Rienzo J.A., Casanoves F., Balzarini M.G., Gonzalez L., Tablada M., Robledo C.W.(2008). InfoStat, versión 2008, Grupo InfoStat, FCA, Universidad Nacional de Córdoba,Argentina.
[36] Llinás, H. (2006). Estadística Inferencial (1.ª ed.). Editorial Universidad del Norte. https://editorial.uninorte.edu.co/gpd-estadistica-inferencial.html.
[37] López, A. (2011). Análisis previo y exploratorio de datos. Sevilla: Universidad de Sevilla. Recuperado de http://personal.us.es/analopez/aed.pdf.
[38] Díaz Chávez, L y Rosado Vega, J. (2019). Tratamiento estadístico de datos con aplicaciones en R. Universidad de la Guajira.