BRNO UNIVERSITY OF TECHNOLOGY

FACULTY OF CIVIL ENGINEERING
institute of building testing

Z 7008

PROCEDURES USED IN THE STATISTICAL ANALYSIS OF PROFICIENCY TESTING PROGRAMS

Proficiency Testing Provider at the SZK FAST
Veveří 95, 602 00 Brno
Czech Republic

szk.fce.vutbr.cz
ptprovider.cz

Coordinator	Assoc. Prof. Ing. Tomáš Vymazal, Ph.D.	Approved: 8/2/2024
Supervisor	Ing. Petr Misák, Ph.D.	Approved: 8/2/2024
Approved for PT Provider	Assoc. Prof. Ing. Tomáš Vymazal, Ph.D.	Approved: 8/2/2024

Contents

1 Procedures used in the Statistical Analysis of Laboratory Results

To describe the accuracy of measuring methods, the terms trueness and precision are used. Trueness refers to the closeness to congruity between the arithmetic mean of a high number of test results and a real or accepted reference value. Precision means the closeness to congruity between test results. The necessity to consider precision is based on the fact that tests generally do not yield the same results even though they are supposed to be carried out on the same material and under the same conditions. This is caused by accidental errors that are impossible to avoid. These errors represent an integral part of every testing procedure and we are unable to control them fully. The comparative analysis of laboratory data does not focus on assessing the trueness of test results, but first and foremost on their precision. Results are thus compared with one another and not with any reference value or real value.

The basis of the statistical analysis is a critical data assessment complying with ISO 5725-2 [1], i.e. the determination of dubious and outlying values, and other irregularities. This assessment is carried out using mainly Grubbs' and Cochran's tests (numerical evaluation) as well as Mandel's statistics (graphical evaluation). Other observed statistical parameters are interlaboratory dispersion, repeatability dispersion, reproducibility dispersion and related characteristics of repeatability and reproducibility. The outcome of PTP is to assess the performance of participating laboratories in compliance with EN ISO/IEC 17043 [2], consisting of the determination of relative values and their uncertainties and a final comparison with the test results of PTP participants.

A prerequisite for using these methods is the unimodal probability distribution of measured data. Furthermore, p will stand for the number of participating laboratories marked by the index A prerequisite for using these methods is the unimodal probability distribution of measured data. Furthermore, p will stand for the number of participating laboratories marked by the index $i = 1, \dots, p$, each of which carried out n number of tests., each of which carried out n number of tests.

1.1 The Numerical Procedure for Determining Outliers

To determine outliers, two basic statistical tests are used. One of them is Cochran's C test, which tests interlaboratory variabilities (in cases when the number of measurements of one quantity in one laboratory > 2) and is used first. If this test marks one participant's results as outlying, the laboratory is excluded and the test repeated. The second test (Grubbs' test) is first and foremost a test of interlaboratory variability and we can also employ it if Cochran's test raises the suspicion that only one of the test results is to blame for the high interlaboratory dispersion. Both tests assume a balanced experiment, i.e. the number of tests at one laboratory for the determination of one quantity must be constant.

When determining divergent or outlying values, three situations can occur:

- If the test statistic is within or equal to 5% of the critical value, the tested entity is considered to be *correct*;
- If the test statistic diverges from the critical value by more than 5%, but is within or equal to 1% of the critical value, the tested entity is considered to be *divergent*;
- If the test statistic diverges from the critical value by more than 1%, the tested entity is considered to be *outlying*.

1.1.1 Cochran's test

The Cochran's C statistic is given by the equation:

$$C = \frac{s_{max}^2}{\sum_{i=1}^{p} s_i^2} \tag{1}$$

where s_{max} is the highest sample standard deviation, s_i are sample standard deviations determined according to the results from all laboratories and p means the number of laboratories participating in the PT program.

The sample standard deviation is determined from the equation

$$s_i = \sqrt{\frac{1}{n_i - 1} \sum_{k=1}^{n_i} (y_k - \bar{y})^2}, \tag{2}$$

where n_i is the number of test results from the determination of one quantity in i-th laboratory, y_k is the k-th value and $\bar{y}_i$ is the average value measured in the i-th laboratory. If only two results were measured for the relevant quantity, we can use the simplified equation:

$$s_i = \frac{|y_1 - y_2|}{\sqrt{2}}. \tag{3}$$

1.1.2 Grubbs' test – One Outlying Observation

From the given set of x_i data for $i = 1, 2, ..., p$, ordered upward according to size, Grubbs' statistic G_p is calculated in order to use Grubbs' test to determine whether the largest observation is an outlier:

$$G_p = \frac{x_p - \bar{x}}{s}, \tag{4}$$

whereby $\bar{x}$ is the arithmetic mean of the observed feature. The observed feature can be the average value of the quantity determined within the laboratory. Furthermore, s is a sample standard deviation of the observed feature, which in this case is a standard deviation calculated for all the laboratories.

For significance testing of the smallest observation the test statistic is calculated:

$$G_p = \frac{\bar{x} - x_p}{s}. \tag{5}$$

1.2 Mandel's Statistics

In order to determine data consistency, two values called Mandel's h and k statistics were used. These indicators are commonly used for the graphical evaluation of laboratories in a similar way to a description of variability.

1.2.1 Interlaboratory Consistency Statistic h

For each laboratory, the interlaboratory consistency statistic h was evaluated according to the formula

$$h_i = \frac{\bar{y}_i - \bar{\bar{y}}}{\sqrt{\frac{1}{p-1} \sum_{i=1}^{p_j} \left(\bar{y}_i - \bar{\bar{y}}\right)}}, \tag{6}$$

where $\bar{y}_i$ is the average value for the i-th laboratory, $\bar{\bar{y}}$ is the arithmetic mean of all values and p is the number of laboratories. The values of the h_i statistics were plotted on graphs.

1.2.2 Intralaboratory Consistency Statistic k

The intralaboratory consistency statistic k is calculated from the equation

$$k_i = \frac{s_i \sqrt{p}}{\sqrt{\sum_{i=1}^{p} s_i^2}}, \tag{7}$$

where s_i is a sample standard deviation of values measured at the i-th laboratory. Just as with h statistics, the k values are plotted on graphs.

Study of the graphs displaying h and k values may indicate that certain laboratories show a significantly different ordering of results than other studied laboratories. This is caused by a permanently large and/or permanently small dispersion of results or extreme averages of results across all levels.

1.3 Calculation of Variances Estimates

After the elimination of outliers (of laboratories), we can proceed to the calculation of basic variability characteristics, i.e. repeatability dispersion, interlaboratory dispersion and reproducibility dispersion. These characteristics are stated in the form of standard deviations, i.e. after extracting the root. It is advantageous when the variability characteristics and the observed quantity are of the same physical dimensions.

1.3.1 Repeatability Variance

$$s_r^2 = \frac{\sum_{i=1}^{p}(n_i - 1)s_i^2}{\sum_{i=1}^{p}(n_i - 1)} \tag{8}$$

1.3.2 Interlaboratory Variance

$$s_L^2 = \frac{s_d^2 - s_r^2}{\bar{\bar{n}}}, \tag{9}$$

where

$$s_d^2 = \frac{1}{p-1}\sum_{i=1}^{p} n_i \left(\bar{y}_i - \bar{\bar{y}}\right)^2 \tag{10}$$

and

$$\bar{\bar{n}} = \frac{1}{p-1}\left[\sum_{i=1}^{p} n_i - \frac{\sum_{i=1}^{p} n_i^2}{\sum_{i=1}^{p} n_i}\right]. \tag{11}$$

1.3.3 Reproducibility Variance

$$s_R^2 = s_r^2 + s_L^2, \tag{12}$$

where s_r^2 is repeatability variance and s_l^2 is interlaboratory variance.

1.4 Repeatability and Reproducibility

Repeatability expresses the fact that the difference between two test results from the same sample from tests carried out by the same person at the same facility and within the shortest time interval possible will not exceed the repeatability value r on average more than once in 20 cases if the method is employed in the common and correct manner.

The repeatability value is expressed by the relation

$$r = 2,8 s_r, \tag{13}$$

where $s_r = \sqrt{s_r^2}$ stands for the standard deviation of repeatability.

Reproducibility expresses the fact that the reproducibility value R for test results from one sample obtained in the shortest time interval possible by two persons who used their own devices will not differ on average more than once in 20 cases if the method is employed in the common and correct manner.

The reproducibility value is expressed by the relation

$$R = 2,8 s_R, \tag{14}$$

where $s_R = \sqrt{s_R^2}$ stands for the standard deviation of reproducibility.

1.5 Assigned Values

The PT Provider will ensure the determination of assigned value X and its uncertainty for every PTP. Assigned values are always only imparted to PTP participants after they have submitted their PTP results so that they cannot obtain any benefit from the premature revelation of the values.

The assigned values are determined by the PT Provider as consensual values derived from the results of participants in compliance with Appendix B of EN ISO/IEC 17043 [2] using the statistical methods described in ISO 13528 [3] and ISO 5725-5 [4]. The assigned value X is therefore determined as a robust estimate of the average value x^* (the A algorithm mentioned in [3] and [4]):

Initial values x^* and s^* (robust standard deviation) are calculated as

$$x^* = \text{median}\ x_i, \tag{15}$$

$$s^* = 1,483 \cdot \text{median}\left|x_i - x^*\right|, \tag{16}$$

where $i = 1, \dots, p$. The values of x^* and s^* are then processed as follows. First, $\varphi = 1,5 \cdot s^*$ is computed. For every x_i $(i = 1, \dots, p)$ value, the following is calculated

$$x_i^* = \begin{cases} x^* - \varphi & \text{if } x_i < x^* - \varphi, \\ x^* + \varphi & \text{if } x_i > x^* + \varphi, \\ x_i & \text{in other cases.} \end{cases} \tag{17}$$

New values of x^* and s^* are calculated from the following equations

$$x^* = \sum_{i=1}^{p} \frac{x_i^*}{p}, \tag{18}$$

$$s^* = 1,134 \cdot \sqrt{\sum_{i=1}^{p} \frac{(x_i^* - x^*)^2}{p-1}}. \tag{19}$$

Robust estimates are derived by iteration until the estimate changes between calculations become small.

The standard uncertainty u_X of an assigned value determined in this manner is calculated from the relation

$$u_X = 1,25 \cdot \frac{s^*}{\sqrt{p}}. \tag{20}$$

In the case of a small number of PTP participants, the PT Provider sets the assigned values as consensual values obtained from expert participants who have proven their competence to determine the measured quantity that is the subject of testing.

Furthermore, if the number of participants is small $(4 \leq p \leq 20)$, the PT Provider can consider determining the relative values by using what is called **Horn's method**. This method consists in the determination of so-called pivots used as a basis for estimating location and variability. First, the assessed data are ordered upwards. The low pivot is then determined from the equation

$$x_D = x_{(H)}, \tag{21}$$

where H is an ordinal index given by the equation $H = \frac{\text{int}\left(\frac{p+1}{2}\right)}{2}$ or $H = \frac{\text{int}\left(\frac{p+1}{2}+1\right)}{2}$.
The upper pivot is then determined from the equation

$$x_H = x_{p+1-H}. \tag{22}$$

Using Horn's method, the assigned value is determined as a location estimate, i.e. as the so-called pivot half sum:

$$x^* = \frac{x_D + x_H}{2}. \tag{23}$$

The variability estimate is determined as the so-called pivot range

$$R_L = x_H - x_D \tag{24}$$

and the uncertainty of an assigned value calculated in this way is determined as a 95% interval estimate of the mean value

$$u_X = R_L \cdot t_{L;0,95}(p), \tag{25}$$

where $t_{L;0,95}(p)$ is the $(1 - \alpha)$ quantile of the T_L probability distribution with p degrees of freedom.

1.6 Calculation of Performance Statistics

Proficiency test results often need to be transformed into performance statistics in order to aid interpretation and to allow comparison with defined objectives. The aim is to express the divergence from the assigned value in a way that enables its comparison with performance criteria. In compliance with the EN ISO/IEC 17043 standard [2], the performance of participating laboratories is evaluated according to the so-called z-score and ζ-score (zeta-score).

For every non-outlying laboratory (participant), the z-score is calculated according to the equation

$$z_i = \frac{|\bar{x_i} - x^*|}{s^*}. \tag{26}$$

ζ-score is calculated using the equation

$$\zeta_i = \frac{|\bar{x_i} - x^*|}{\sqrt{u_i^2 + u_X^2}}, \tag{27}$$

where u_i is a combined standard uncertainty of the i-th laboratory. Combined standard measurement uncertainties can be arrived at by dividing the extended uncertainty U by the extension coefficient k, which for normal probability division has the value $k = 2$. If the participant does not state the extended measurement uncertainty in their test result protocol, it is impossible to determine the ζ-score. For more about measurement uncertainties see document [5].

The following scales are applied for the z-score and ζ-score (to simplify the matter, only the z-score is shown):

$$z\text{-score} = \begin{cases} |z| \leq 2 & \text{shows that the laboratory performance is } \textbf{satisfactory} \text{ and generates no signal;} \\ 2 \leq |z| \leq 3 & \text{shows that the laboratory performance is } \textbf{questionable} \text{ and generates an action signal;} \\ 3 \leq |z| & \text{shows that the laboratory performance is } \textbf{unsatisfactory} \text{ and generates an action signal.} \end{cases} \tag{28}$$

References

[1] ISO 5725-2. *Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*. 2019.

[2] EN ISO/IEC 17043. *Conformity assessment - General requirements for proficiency testing*. 2010.

[3] ISO 13 528. *Statistical methods for use in proficiency testing by interlaboratory comparisons*. 2022.

[4] ISO 5725-5. *Accuracy (trueness and precision) of measurement methods and results - Part 5: Alternative methods for the determination of the precision of a standard measurement method*. 1999.

[5] EA 4/02. *Vyjadřování nejistot měření při kalibracích*. 2000.