

1 Methods of the Results Statistical Evaluation PTs MPZ UKZUZ

The MPZ UKZUZ proficiency testing schemes are organized as continuous, regular PT schemes with a specific interval or in defined PT proficiency testing cycles with planned timeframe. They are designed in accordance with customer needs to ensure regular external verification of results in analytical laboratories or for the purpose of verification/validation of analytical procedures.

All procedures of the statistical evaluation of the PTs MPZ UKZUZ results are robust statistical methods. Robust statistical methods were selected for the statistical processing of MPZ UKZUZ PT results to eliminate the influence of outliers on the determination of the assigned value.

PT department NRL UKZUZ (PT provider) uses strictly statistically derived methods for evaluation of participant performance.

Evaluation of performance and statistical calculation are based on robust statistics of the set of submitted results. The assigned value and the standard deviation of individual comparisons are derived from the results submitted by MPZ participants.

The assigned value and the standard deviation value (uncertainty of assigned values) of the evaluation of parameters and characteristics of PT items are determined based on a consensus of the results submitted by PT participants and the variance of their results.

The robust analysis according to algorithm A with iteration - estimation of mean value and standard deviation of data is primarily used for the statistical processing and evaluation of participant's results. The principle of the calculation of z-score are values of the robust average x^* and the robust standard deviation s^* calculated according to algorithm A for the robust analysis. The detailed description can be found in standard ISO 13528 [1].

In case of problematic results, it is possible to use the values determined by the evaluation of homogeneity or stability tests as the assigned values (ATV - assigned target value and s_{ATV} - assigned target value of assigned standard deviation).

For a smaller number of laboratories (4-7), the assigned value and confidence interval are calculated using the Horn procedure [2,4].

If the set of comparison results is very tight for some parameters, the criteria for the statistical elimination of outliers and the subsequent calculation of z-score can be very strict. In this case there can be laboratories that were eliminated by the statistical evaluation although they fulfil the conditions of reproducibility given by the legislation or by standards. An internal evaluation can then be performed at the laboratory based on the reproducibility values specified in the standards.

1.1 Robust statistics calculated according to algorithm A

The calculation of z-score on the principle of robust statistics according to algorithm A is possible only for 8 reported values at least. Z-score calculated according to algorithm A is defined as the difference between the reported value and the value of the robust average x^* divided by the robust standard deviation s^* . Within proficiency testing scheme the absolute value of z-score lower or equal to 1 is considered to be good, the absolute value from 1 to 2

means that the performance is satisfactory, the absolute value between 2 and 3 means that measurements are of questionable quality (in results marked with „!“ and the absolute value higher than 3 means that measurements are out of satisfactory range (in results marked with „!!“). Z-score is calculated and expressed in colour both in the numerical and the graphical forms (absolute value of z-score 0-2 is green, z-score 2-3 is orange and z-score higher than 3 is red).

The calculation of z-score is not carried out when less than 4 results from participants are provided and when the robust standard deviation is calculated as zero ($s^* = 0$).

The calculation of z-score does not include values signed “less than” (“<”). A sign “less than” (“<”) means that a laboratory delivered a result in the form of < detection limit. For the results reported as “<”, the value of limit of quantification (LOQ) is crucial. These values are evaluated according to further criteria. When the result of the laboratory is lower than the difference between robust average and 3fold of robust standard deviation ($x^* - 3 s^*$), the result is marked with „!!“ as outlier. When the result of the laboratory is higher than the sum of robust average and 3fold of robust standard deviation ($x^* + 3 s^*$), the result is marked with „M!“, the laboratory expressed the high value of LOQ.

Calculation of z-score according to algorithm A:

For all analytical data z-score is calculated according to the formula

$$z\text{-score} = (x - x^*) / s^*$$

where:

x – participant’s reported value

x^* – robust average, may be replaced by ATV - assigned target value

s^* – robust standard deviation, may be replaced by s_{ATV} - assigned target value of assigned standard deviation

The following algorithm for the robust analysis [1] is used for the calculation of the robust values of the average x^* and standard deviation s^* :

Denote the p items of data (reported values), sorted into increasing order, by:

$$x_1, x_2, \dots, x_i, \dots, x_p$$

Calculate initial value for x^* , which is median of reported values and calculate initial value of robust standard deviation s^* , which is median of absolute deviations multiplied by factor 1,483:

$$x^* = \text{median of } x_i, \text{ where } i = 1, 2, \dots, p$$

$$s^* = 1,483 \times \text{median of } |x_i - x^*|, \text{ where } i = 1, 2, \dots, p.$$

Update the values of x^* and s^* as follows. Calculate the value of δ according to formula

$$\delta = 1,5 s^*$$

and then for each x_i ($i = 1, 2, \dots, p$), calculate x_i^* , for which one of the following conditions holds true:

$$x_i^* = \begin{cases} x^* - \delta, & \text{if } x_i < (x^* - \delta) \\ x^* + \delta, & \text{if } x_i > (x^* + \delta) \\ x_i & \text{otherwise} \end{cases}$$

From the all values x_i^* ($i = 1, 2, \dots, p$), calculate the new values of x^* and s^* from:

$$x^* = \sum_{i=1}^p x_i^* / p$$

$$s^* = 1,134 \sqrt{\sum_{i=1}^p (x_i^* - x^*)^2 / (p - 1)}$$

The robust estimates x^* and s^* are derived by an iterative calculation, i.e. by updating the values of x^* and s^* several times using the modified data (always from the step of calculating new value of δ using the updated value of s^*), until the process converges. Convergence may be assumed if there is no change from one iteration to the next in the third significant figure of the robust standard deviation.

1.2 Robust statistics calculated according to Horn procedure

This procedure is recommended for the small sample sizes; it can be applied for 4 – 7 valid measurements (reported values). Statistical evaluation is based on the n -order statistics of the set. The depth of pivots, the midpivot (the estimate of location) and the pivot spread (the estimate of the spread of the data) are determined according to the number of reported values n . The product of pivot spread and quantile of distribution T_L for $1 - \alpha = 0,975$ gives an estimation of the confidence intervals.

The pivot depth h is determined according to the relations

$$h = (\text{int} [(n + 1)/2]) / 2 \text{ or } h = (\text{int} [(n + 1)/2] + 1) / 2$$

according to, which number will be the integer. The lower pivot is $x_D = x_{(h)}$ and the upper pivot is $x_H = x_{(n+1-h)}$, n being the number of values.

The estimate of location – the midpivot P_L is determined from the relation

$$x_{HORN} = P_L = (x_D + x_H) / 2$$

and the pivot spread $R_L = x_H - x_D$ is the estimate of the spread of the data.

The distribution of the random variable $T_L = P_L / R_L$ is approximately symmetrical.

The estimate of the confidence interval s_{HORN} is given by the product of pivot spread R_L and quantile of distribution T_L for $1 - \alpha = 0,975$, i.e.

$$s_{HORN} = R_L \times t_{L,0,975}(n)$$

Quantiles $t_L, 1 - \alpha = 0,975$ (n) for distribution T_L see [2].

Quantiles $t_{L,0,975}(n)$ of the T_L distribution for statistical calculation according to Horn's method for the analysis of the small sample sizes

n	$t_{L,0,975}$
4	0,738
5	2,094
6	1,035
7	0,720
8	0,564
9	0,915
10	0,668
11	0,545
12	0,483
13	0,608
14	0,525
15	0,466
16	0,435
17	0,502
18	0,451
19	0,423
20	0,397

Calculation of z-score according to Horn procedure:

Z-score calculated according to Horn procedure (4-7 valid measurements) is defined as the difference between the reported value and the mean value of data set x_{HORN} (estimation of location) divided by the standard deviation s_{HORN} (estimation of confidence interval).

$$z\text{-score} = (x - x_{HORN}) / s_{HORN}$$

The calculated z-score given lower or equal to 1 is considered to be good, the absolute value from 1 to 2 means that the performance is satisfactory, the absolute value between 2 and 3 means that measurements are of questionable quality (in results marked with „!““) and the absolute value higher than 3 means that measurements are out of satisfactory range (in results marked with „!!““).

1.3 Statistical evaluation of PT scheme MPZ UKZUZ Analysis of Mycotoxins

All procedures of the statistical evaluation of the PT MPZ UKZUZ – Analysis of Mycotoxins results are robust so that the influence of outliers on an assessment of the assigned values would be excluded. Z-score is calculated and expressed both in numerical and graphical forms.

The assigned value is calculated for each parameter of MPZ UKZUZ – Analysis of Mycotoxins according to the algorithm for the robust analysis in accordance with ISO 13528 [1] as the robust average x^* of the values received from participants who carried out their analyses by the chromatographic methods.

In the case of smaller number of participants (4-7) the value of z-score is calculated according to Horn procedure for small sample sizes [4] or value ATV (assigned true value) obtained during homogeneity test is used for z-score calculation.

The value of the standard deviation for the calculated assigned value (from the results received by the chromatographic methods) is calculated according to the Horwitz equation, actually equations modified according to Thompson [3], depending on the analyte concentration expressed as mass fraction.

$$s_{Horwitz} = \begin{cases} 0,22 c & \text{when } c < 1,2 \times 10^{-7} \\ 0,02 c^{0,8495} & \text{when } 1,2 \times 10^{-7} \leq c \leq 0,138 \\ 0,01 c^{0,5} & \text{when } c > 0,138 \end{cases}$$

$s_{Horwitz}$ – standard deviation of reproducibility calculated according to the modified Horwitz/Thompson equation

c – the concentration of the analyte/parameter (mass fraction in the format of unit/unit e.g. g/g or µg/µg). In fact it is the robust average x^* calculated according to ISO 13528 [1], expressed as the mass fraction in the required format (it can be replaced by ATV – assigned true value i.e. the value, which is determined in advance or by the value of x_{Horn} calculated according to the Horn procedure for 4-7 number of participants).

The results from diagnostic methods (ELISA, ROSA) are processed separately using assigned value obtained by statistical processing of result from chromatographic methods (entered into z-score calculation as ATV) and the standard deviation of reproducibility calculated according to the modified Horwitz equation (entered into the z-score as s_{ATV}).

References:

1. ISO 13528 Statistical methods for use in proficiency testing by interlaboratory comparison, 2022, Annex C.
2. Meloun, M., Militký, J.: Komentář k statistickému zpracování dat. Academia, Praha, 2002, str.146-147.
3. Thompson M.: Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing, Analyst, 2000, 125, 385–386
4. Horn, P. S. „Some Easy Statistics“. Journal of the American Statistical Association, December 1983, Volume 78, Number 384, Theory and Methods Section.