FNRI PROFICIENCY TESTING SUPPLEMENT ON STATISTICAL PROCEDURES

November 2022
7th Edition

JOLLY C. COTARA
Technical Manager

LEAH C. DAJAY
PTL Head

DOST-FNRI

Department of Science and Technology
Food and Nutrition Research Institute
PROFICIENCY TESTING LABORATORY

PREFACE

This document describes the statistical procedures that the Department of Science and Technology – Food and Nutrition Research Institute (DOST-FNRI) uses in (a) the analysis of the results of its proficiency testing (PT) programs, (b) evaluation of test material homogeneity, and (c) evaluation of test material stability.

The procedures for the evaluation of PT results and test material homogeneity and stability detailed here are based on ISO 13528:2022 and the International Harmonized Protocol for the Proficiency Testing of Analytical Chemical Laboratories (IUPAC Technical Report, 2006). For procedures, which are not specified in the aforementioned documents, the best judgment of the author was relied on.

Other statistical procedures and plots detailed in ISO 13528:2022 and IUPAC Protocol may be applied, whenever necessary and applicable, and with proper guidance from a statistician.

I. EVALUATION OF PROFICIENCY TESTING (PT) RESULTS

A. INTRODUCTION

Proficiency test results are assessed by comparison with assigned values derived from the consensus of results (consensus value) from participants, or values determined by a reference laboratory.

The consensus values are estimated using robust procedures. Robust procedures are used in the estimation of consensus values because the most commonly used measures of location and dispersion – **arithmetic mean and standard deviation** – are highly influenced by the presence of extreme outliers and their interpretation depends on an implicit assumption that they are a random sample from a normal distribution. The mean and standard deviation are the optimal estimators of location and dispersion, respectively, for a normal distribution but they can be substantially sub-optimal for distributions close to the normal.

It is very common in many fields to encounter data that have skewed distributions or contain outliers. Analytical data from testing laboratories often depart from the assumption that the data are a random sample from a normal distribution. It is often heavy tailed – contains a higher than expected proportion of results far from the mean – and sometimes contains outliers.

Outliers are values that are so far in value from the rest of the data that they may be viewed as coming from a different population, or the result of a measurement error. One way of coping with outliers is to exclude them from the calculation of the statistics. But when is it justifiable to exclude outliers in the calculation? The decision to exclude or retain an outlier depends on the understanding of the cause of the outlier and its impact on the results.

On the basis of some simple assumptions, outlier tests identify where it is likely to have a technical error but it does not assess or judge that the point is "wrong". In a data set, the value may be extreme but it could be the correct one. Only with experience or by identification of a certain cause can data be declared "wrong" and excluded from the computations. Generally, if more than 20% of the data are identified as outlying, the assumption about the data distribution and/or the quality of the data collected becomes questionable.

A convenient way of coping with outliers is to use **robust statistics**. Robust statistics includes methods that are largely unaffected by the presence of extreme values. "It provides an alternative way of summarizing results when they include a small proportion of outliers, without the requirement to identify specific observations as outliers or exclude them" [**1**].

Examples of robust statistics are the median and the mode for they are not highly influenced by the presence of outliers. "The ***median*** is the value in an ordered data set that has an equal number of data points on either side while the ***mode*** is the value of the peak of the distribution" [**2**].

Among the three statistics – mean, median and mode – the mode is least affected by the presence of outliers. However, because the calculation of the mode is more difficult than that of the mean or median, the mode has limited application.

B. SETTING THE STANDARD DEVIATION FOR PROFICIENCY ASSESSMENT

"The standard deviation for proficiency assessment (σ_{pt}) is a parameter that is used to provide a scaling for the laboratory deviations from the assigned value and thereby define a *z* score. The value is determined by "fitness-for-purpose" as it does not represent a general idea of how laboratories are performing, but how they ought to perform to fulfill their commitment to their clients" [**3**].

Fitness-for-purpose is the ability of a value to satisfy a set of conditions given by the application. "The **uncertainty of measurement** is a parameter associated with the results of a measurement that characterizes the dispersion of the value that could reasonably be attributed to the measurand" [**4**].

Most common approaches in setting the σ_{pt} of a measurand are the following:

(i) by collaborative trial data calculated using the formula:

$$\sigma_{pt} = RSD_R \times x_{pt}$$

where:

RSD_R is the relative standard deviation of reproducibility from collaborative trials

x_{pt} is the assigned value from consensus of PT participants' results derived as a robust average using Algorithm A of ISO 13528 expressed in appropriate units

(ii) by perception of how laboratories should perform, based on CV of previous PT results on appropriate (same or similar) matrix

$$\sigma_{pt} = \frac{(CV \times x_{pt})}{100}$$

where:

CV is the coefficient of variation

x_{pt} is the assigned value from consensus of PT participants' results, derived as a robust average using Algorithm A of ISO 13528 expressed in appropriate units

(iii) by perception of the Technical Working Group (TWG) on how PT participants should perform

(iv) by experience from previous PT Rounds for the same measurand with comparable property values, and where PT participants use compatible measurement procedures

(v) from the uncertainty of the certified/reference value from CRM/SRM, computed as follows:

$SD = U/k$ $\quad\quad$ $\sigma_{pt} = SD$

where:

U	is the uncertainty of the certified/reference value of the CRM/SRM
SD	is the standard deviation
K	is the coverage factor

(vi) by Horwitz equation in the absence of collaborative trial data for minerals using the formula:

$\sigma_{pt} = 0.02 \times c^{0.8495}$

where:

c	is the concentration (i.e., assigned value) of the measurand in mass fraction, i.e., when $1.2 \times 10^{-7} \leq c \leq 0.138$

Note: *Not applicable for macro-analysis, e.g., proximates*

(vii) by using the robust standard deviation (s*), computed based on ISO 13528:2022 Algorithm A or robust CV of participants,

σ_{pt} = robust standard deviation, s*

or

σ_{pt} = (robust CV x robust average) / 100

Note:

1. Robust CV of participants from previous similar rounds or current round (whichever is lower) may be used by the PT provider

2. When $\sigma_{pt} = s^*$, re-compute or verify homogeneity and stability test results using s.*

(viii) by using the predicted relative standard deviation of reproducibility ($PRSD_R$) set in the AOAC Official Methods of Analysis (2019):

Analyte	RSD_R
100%	2%
10%	3%
1%	4%
0.1%	6%
100 ppm (mg/kg)	8%
10 ppm (mg/kg)	11%
1 ppm (ppm)	16%
100 ppb (ug/kg)	22%
10 ppb (ug/kg)	32%

$\sigma_{pt} = (RSD_R \times C) / 100$

(ix) by using the reproducibility precision relative standard deviation, RSD_i set by the FNRI- Service Laboratory based on the method validation data:

1. RSD_i for proximates is $\leq$ 5%;
2. RSD_i for minerals is $\leq$ 10% or $\leq$ 15% if value is < 5 mg for Na, Ca, K and value is < 0.5 mg for Fe and Zn; and
3. RSD_i for other nutrients is $\leq$ 10 %

$$\sigma_{pt} = (RSD_i \times x^*) / 100$$

where:
RSD_i is the reproducibility precision
x^* is the robust average

The standard deviation of reproducibility found in collaborative trials is generally considered an appropriate indicator of the best agreement that can be obtained between laboratories [**5**].

C. EVALUATION PROCESS FOR PT RESULTS

The evaluation of proficiency test results proceeds as follows:

- **Exclusion of invalid data**

  There may be instances where a participant's test result will be excluded from the calculation of a measurand's consensus value and its associated standard uncertainty. Reasons for exclusion are the following:

  a. method used for the measurand is not applicable to the food matrix (e.g., fat analysis using direct solvent extraction instead of acid/alkaline hydrolysis); and
  b. removal of extreme results or results that are identifiably invalid, (e.g., results caused by calculation errors or used wrong unit of measurements).

- **Determination of the assigned value (x_{pt}) and its standard uncertainty ($u(x_{pt})$)**

  - Calculation of the robust average (x^*) for use as consensus value and the corresponding robust standard deviation (s^*) of the test results;
  - Computation of the standard uncertainty of the consensus value ($u(x_{pt})$) using the formula:

$$u(x_{pt}) = \frac{1.25 \times s^*}{\sqrt{p}}$$

where:

s^*	is the robust standard deviation computed using Algorithm A of ISO 13528:2022
p	is the number of data included in the computation of consensus value

- **Calculation of performance statistics**

  z scores are typically used in the evaluation of performance. The z scores are calculated, using the consensus value and σ_{pt}, only when the consensus value is suitable for use as an assigned value. Otherwise, z' scores is issued in the evaluation of laboratory performance.

  Section I.C.2 gives the details on the calculation of performance statistics.

- **Evaluation of performance**

  Section I.C.3 describes the steps in evaluating the performance of participating laboratories.

Graphical representations of some evaluation outputs are in Figures 1 to 5. The flowchart of the statistical evaluation process is presented in Figure 5 and the detailed steps and explanatory notes are given below.

1. Determination of the Assigned Value (x_{pt}) and its Standard Uncertainty ($u(x_{pt})$)

1.1. *Construction of a kernel density plot of the results*

Kernel density plot is constructed to describe the general shape of the distribution of the data set. The construction of kernel density plot is done by a Kernel add-in software in MS Excel, which is downloaded from AMC Software of Royal Society of Chemistry website.

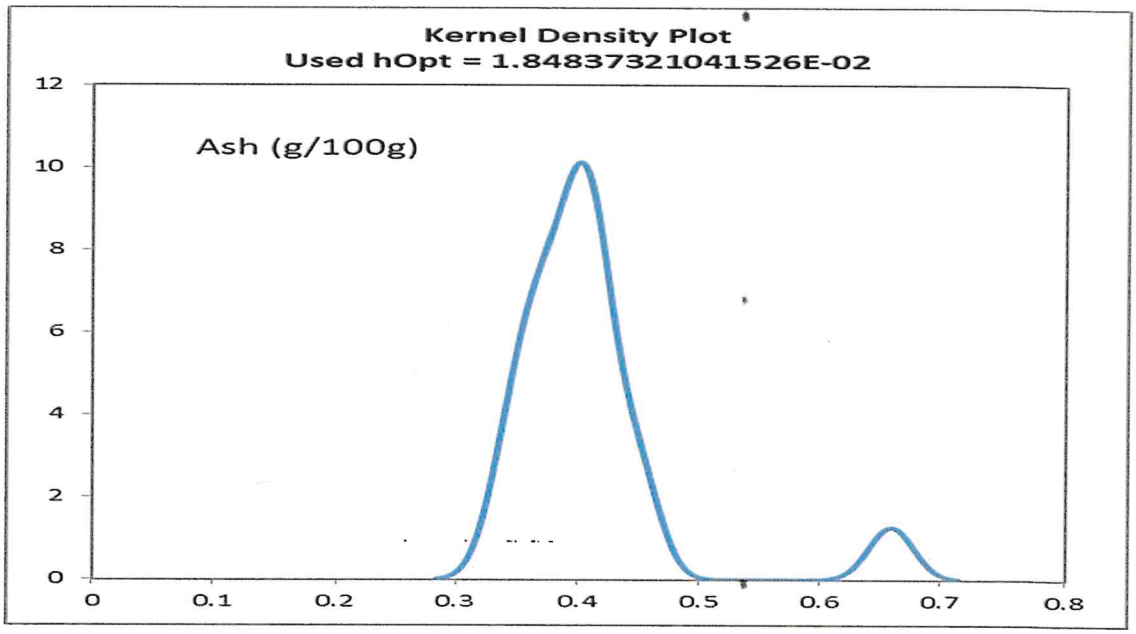


Figure 1. Sample of a Kernel density plot

1.2. *Exclusion of data obtained using inapplicable method or expressed in wrong units of measure (e.g. fat analysis using direct solvent extraction instead of acid/alkaline hydrolysis prior to solvent extraction)*

1.3. *Calculation of the robust average (x*) for use as consensus value and the corresponding robust standard deviation (s*) of the test results using Algorithm A of ISO 13528:2022*

1.4. *Calculation of the standard uncertainty of the consensus value ($u(x_{pt})$) using the following formula:*

$$u(x_{pt}) = \frac{1.25 \times s^*}{\sqrt{p}}$$

where:

p is the number of data included in the computation of the robust average, and

s^* is the robust standard deviation computed using Algorithm A

1.5. *Determination of the suitability of the consensus value to be used as assigned value based on the ISO 13528:2022 criteria [3]:*

if $u(x_{pt}) \leq 0.3\sigma_{pt}$ - $u(x_{pt})$ is negligible, z scores can be issued;

if $u(x_{pt}) > 0.3\sigma_{pt}$ - $u(x_{pt})$ is high, use the uncertainty of the assigned value in the interpretation of performance, i.e., z' scores can be issued if $u^2(x_{pt}) + \sigma^2_{pt} \leq s^{*2}$ (wherein σ_{pt} is not derived from robust CV of participants and $s^* > \sigma_{pt}$)

where:

$u(x_{pt})$ is the standard uncertainty of the assigned value

σ_{pt} is the standard deviation for proficiency assessment
s^* is the robust standard deviation

1.6 *Abandonment of attempt to determine a consensus value*

The attempt to determine a consensus value is abandoned if the uncertainty of the consensus value is not negligible or is too high and if $u^2(x_{pt}) + \sigma^2_{pt} \leq s^{*2}$ is not met (provided that σ_{pt} is not derived from robust CV of participants and $s^* > \sigma_{pt}$). There is no real consensus of results, thus no z or z' score is issued. However, the participants are provided with summary statistics (e.g. mean, median) of the data set as a whole.

2. Calculation of Performance Statistics

z scores are the basis for evaluating the performance of participating laboratories. The z scores are calculated using the consensus value and the standard deviation for proficiency assessment (σ_{pt}), only when the consensus value is suitable for use as the assigned value. The performance of individual PT participant laboratories was evaluated using the formula:

$$z = \frac{x - x_{pt}}{\sigma_{pt}}$$

where
x is the participant's reported result
x_{pt} is the assigned value from the consensus of the PT participants' results derived as robust average
σ_{pt} is the standard deviation for proficiency assessment

The laboratory z scores are interpreted as follows:

|z score| ≤ 2.00: "Satisfactory" **(S)** performance
2.00 < |z score| < 3.00: "Warning" **(W)** signal
|z score| ≥ 3.00: "Action" **(A)** signal

If the uncertainty of the assigned value is greater than $0.3\sigma_{pt}$, then the uncertainty can be taken into account by expanding the denominator for the calculation of performance score, such that:

$$z' = \frac{x - x_{pt}}{\sqrt{\sigma^2_{pt} + u^2(x_{pt})}}$$

where:
x is the participant's reported result
x_{pt} is the assigned value from the consensus of PT participants' result

σ_{pt} is the standard deviation for proficiency assessment
$u(x_{pt})$ is the standard uncertainty of the assigned value

z' scores are interpreted in the same way as *z* scores and using the same critical values of 2.00 and 3.00.

The plots of ordered test results with expanded uncertainty, ordered test results according to methods used and ordered *z* or *z'* scores are also used in evaluating performance. These are graphical means by which a participating laboratory can readily compare its performance relative to the other laboratories.

3. Construction of Plots

3.1 *Construction of plot of ordered test results with expanded uncertainty*

The plot of ordered test results with expanded uncertainty (Figure 2) is a graphical display of each laboratory's test result with the reported expanded uncertainty. It shows the performance of each laboratory relative to the other laboratories. For example, in Figure 2, the test results starting from **Lab 024** to **Lab 025** are within the range of values of "Satisfactory" range: 20.37 to 23.06 g/100g. However, the test results of **Labs 029, 038, 007** and **037** are below the lower limit of the value for "Satisfactory" range, while **Labs 016** and **021** obtained test results above the upper limit of the value for "Satisfactory" range.

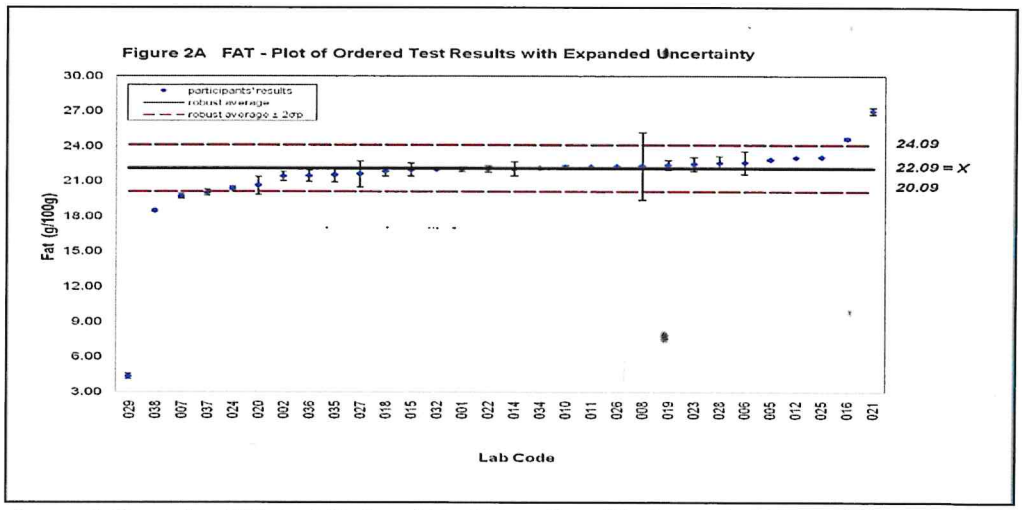


Figure 2. Sample of Plot of Ordered Test Results with Expanded Uncertainty

3.2 *Construction of plot of ordered test results according to methods used*

This plot is a graphical display of the participant's performance according to methods used. It shows if there are differences and clustering of results by method used. For example in Figure 3, comparable behavior of results was observed for both alkali and acid hydrolysis for fat (i.e. no clustering of data).

When there is clustering of results, there is a need to conduct separate evaluation of participants' results by method used.

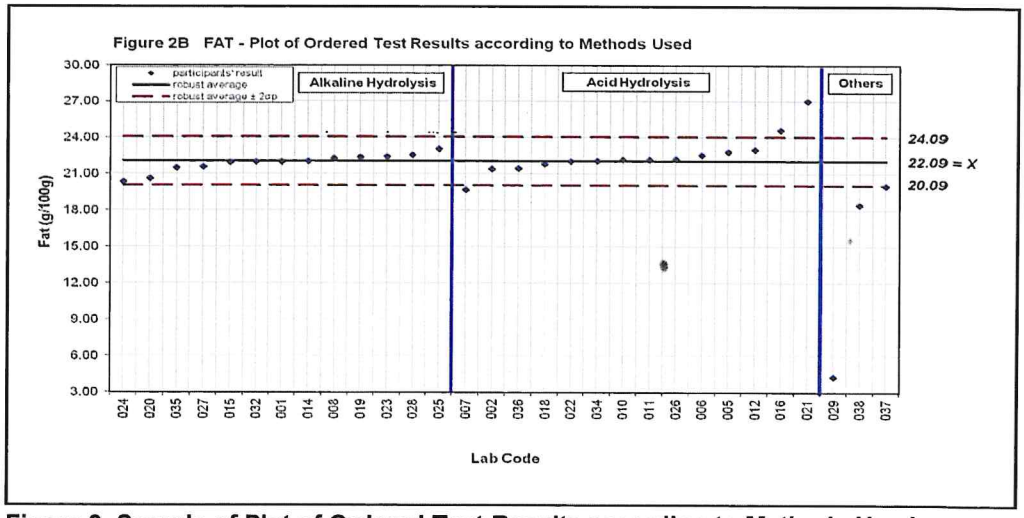


Figure 3. Sample of Plot of Ordered Test Results according to Methods Used

3.3 *Construction of plot of ordered z or z' scores*

The plot of ordered *z* or *z'* scores is a graphical display of the participants' performance. This plot shows each participant laboratory's performance relative to that of the other laboratories. From this plot, results outside the "Satisfactory" range (i.e. (| *z* or *z'* score | > 2.00) can be quickly identified. As illustrated in Figure 4, **Labs 027, 002** and **001** have results outside the "Satisfactory" range.

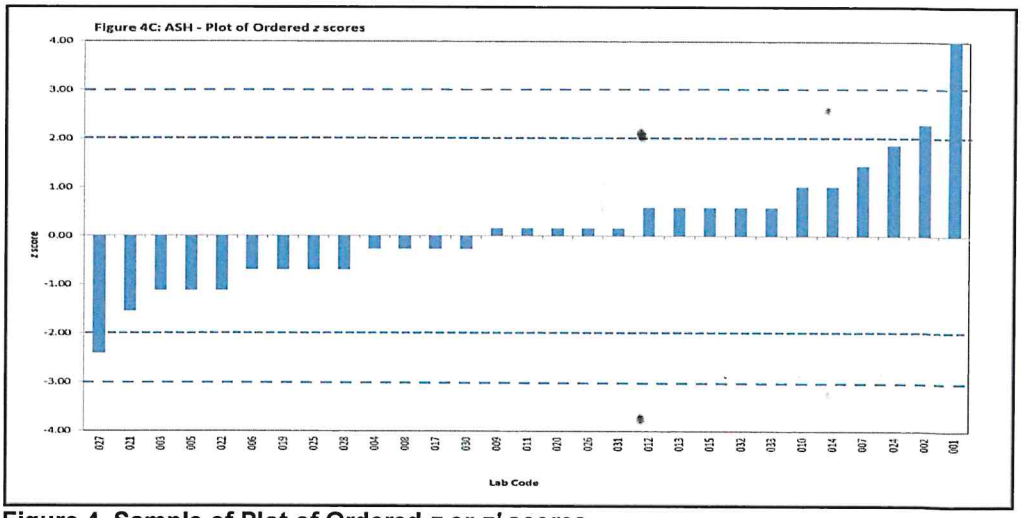
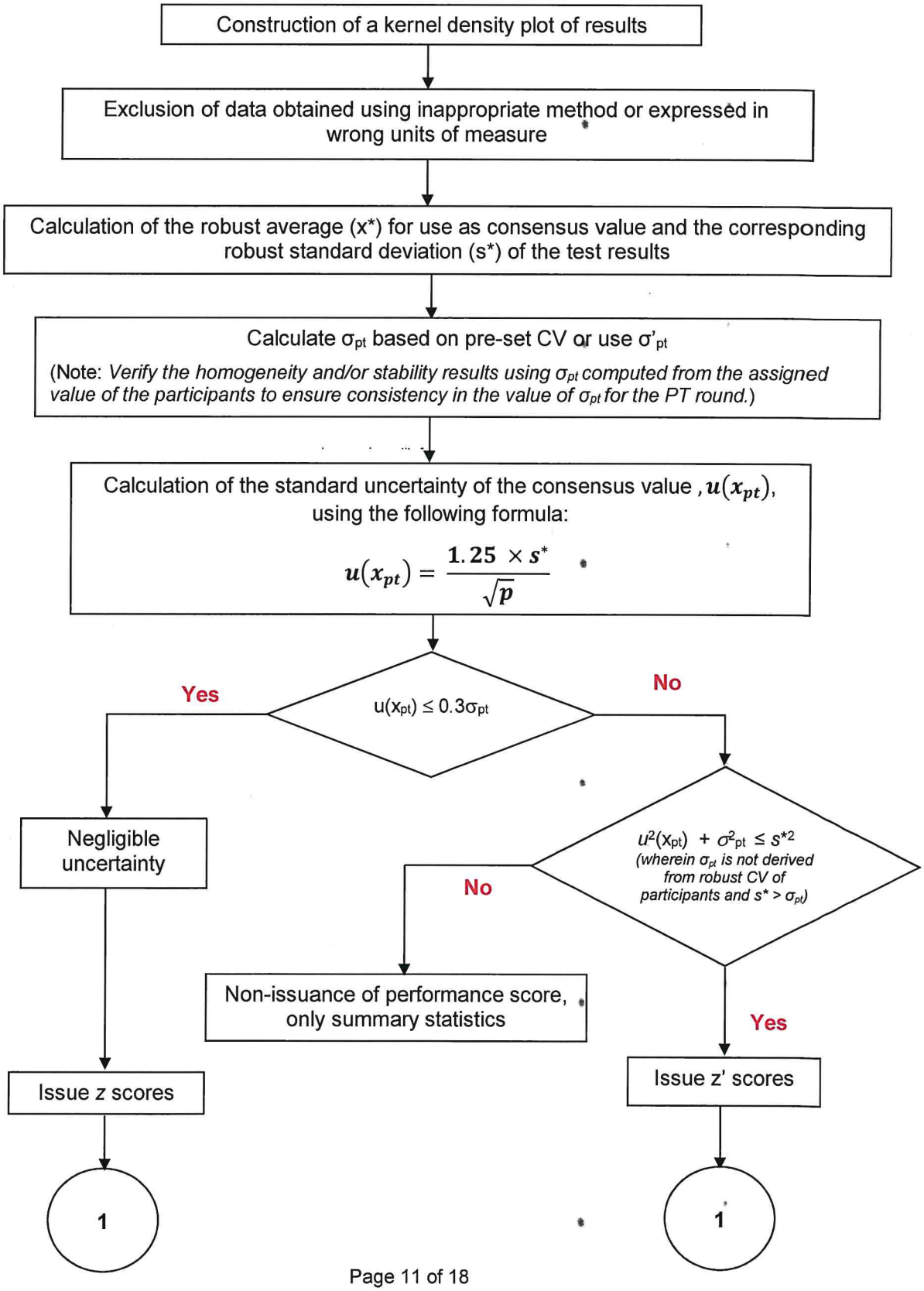
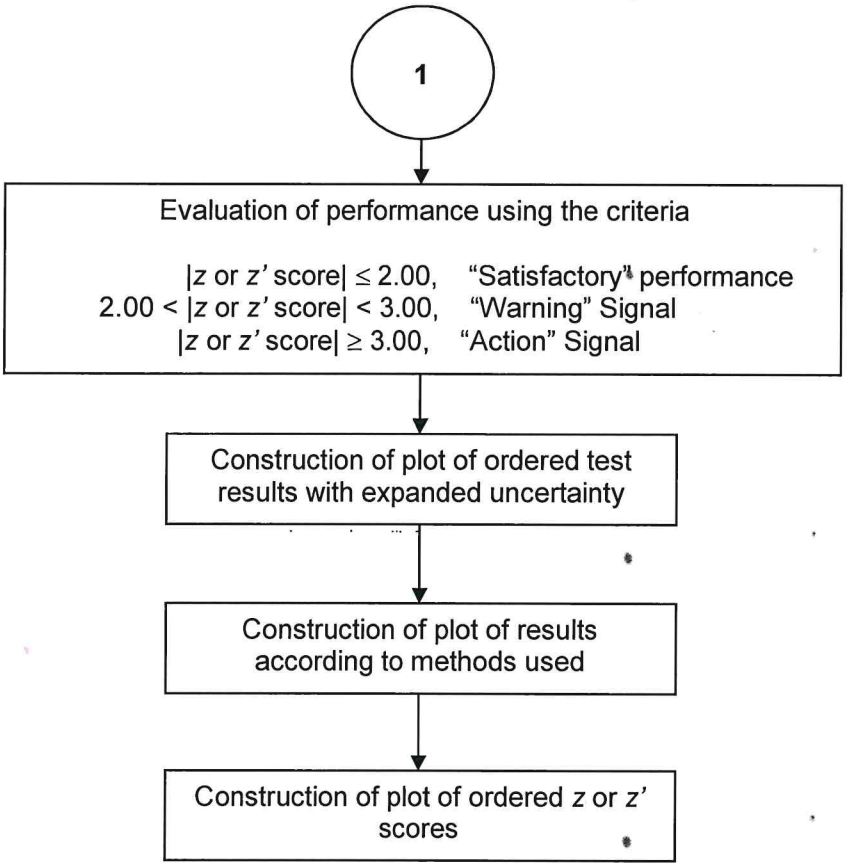


Figure 4. Sample of Plot of Ordered *z* or *z'* scores

Figure 5. STATISTICAL EVALUATION PROCESS




1
Evaluation of performance using the criteria
|z or z' score| ≤ 2.00, "Satisfactory" performance
2.00 < |z or z' score| < 3.00, "Warning" Signal
|z or z' score| ≥ 3.00, "Action" Signal
Construction of plot of ordered test results with expanded uncertainty
Construction of plot of results according to methods used
Construction of plot of ordered z or z' scores

II. EVALUATION OF TEST MATERIAL HOMOGENEITY

The statistical methods used in testing a material for homogeneity are:

- Cochran's test procedure for duplicate results
- Test for "adequate" homogeneity using ISO 13528 assessment criterion for homogeneity check
- Test for "sufficient" homogeneity based on ISO 13528 criterion

The following are the steps in conducting the homogeneity test:

1. **Selection** of 10 test samples in their final packaged form using systematic sampling using Microsoft Excel software,

2. **Separate homogenization** of the contents of each of the 10 selected packages by the appropriate techniques, to minimize within-package variability,

3. **Preparation** of two (2) sub-samples from each test sample using techniques appropriate to the test material, to minimize between-test-portion differences,

4. **Obtaining of a measurement result** on each of the twenty (20) sub-samples in a random order as in Step 1 of this Section, where applicable, and completing the whole series of measurements under repeatability conditions (i.e. same laboratory, same analyst, same method and equipment),

5. **Examination of data** for pathologies,

   5.1. *Construction of a simple plot of duplicate results*

   Use any software that has the capability to construct a scatterplot of duplicate results, e.g., Excel.

   5.2. *Visual examination of a simple plot of the duplicate results and searching for diagnostic features such as:*

   - trends or discontinuities
   - nonrandom distribution of differences between first and second test results
   - excessive rounding; and
   - outlying results within samples

   5.3. *Testing for method precision*

   5.3.1. Calculation of within-sample variance, s_w^2

$$s_w^2 = \frac{1}{g}\sum_{t=1}^{g} s_t^2$$

where:
s_t^2 is the between-test-portion variance
g is the total number of samples, i.e., 10

5.3.2. Computation for the within-sample standard deviation, s_w, or the repeatability standard deviation, s_r

$$s_r = s_w = \sqrt{s_w^2}$$

5.3.3. Use of the ratio of the calculated repeatability standard deviation, s_r, to σ_{pt} in making decision and comparing it with the critical value for method precision which is 0.5

5.3.3.1. If $s_r / \sigma_{pt} \geq 0.5$, the test for method precision has failed.

5.3.3.2. If $s_r / \sigma_{pt} < 0.5$, the test for method precision has been passed. The method is precise enough to detect significant inhomogeneity.

5.4. *Testing for repeatability outliers or outlying results within samples using Cochran's test procedure for duplicate results*

The Cochran's test procedure is as follows:

5.4.1. Calculation of the sum, S_i, and difference, D_i, of each pair of duplicates, for i = 1, ..., g, where g = 10

5.4.2. Calculation of the sum of squares, S_{DD}, of the 10 differences

$$S_{DD} = \sum D_i^2$$

5.4.3. Calculation of the ratio, C, and comparison of the result with the appropriate critical value

The Cochran's test statistic is the ratio of D^2_{max}, the largest squared difference, to this sum of squared differences

$$C = D^2_{max}/S_{DD}$$

For 10 test samples analyzed in duplicate, the critical values at 95% and 99% levels of confidence are 0.602 and 0.718, respectively. Refer to the IUPAC Technical Report for other values.

5.4.3.1. Close inspection of outlying pairs detected at the 95% or higher level of confidence for transcription or other errors. An outlying pair is rejected when there are irremediable analytical errors or if the difference between duplicate results is significant at the 99% level.

5.4.3.2. Deletion of duplicate results from a single test sample if they are significantly different from each other at the 99% level of significance.

5.4.3.3. Discarding of data if they contain discrepancies in two or more test samples. However, pairs of results with outlying

mean (average) value but with no evidence of extreme variance (difference) are not discarded.

6. Testing for homogeneity

6.1. *Testing for "adequate" homogeneity*

6.1.1. Calculation of the variance of the sample averages, $s_{\bar{x}}^2$

$$s_{\bar{x}}^2 = \frac{\sum(\bar{x}_t - \bar{\bar{x}})^2}{(g-1)}$$

where:

$\bar{x}_t$ are the sample averages
$\bar{\bar{x}}$ is the general average
g is the total number of samples, i.e. 10

6.1.2. Calculation of the between-sample standard deviation, s_s

$$s_s = \sqrt{\max\left(0, s_{\bar{x}}^2 - \frac{s_w^2}{2}\right)}$$

Note: The estimate of between-sample variance s_s^2 often becomes negative when s_s is relatively smaller than s_w. This can be expected when test items are highly homogeneous. In this case, $s_s = 0$.

6.1.3. Comparing the between-sample standard deviation, s_s, with $0.3\sigma_{pt}$

6.1.3.1. If $s_s > 0.3\sigma_{pt}$, the test for adequate homogeneity has failed.

6.1.3.2. If $s_s \leq 0.3\sigma_{pt}$, the test for adequate homogeneity has been passed.

6.2. *Testing for "sufficient" homogeneity*

6.2.1. Calculation of the allowable sampling variance, σ^2_{allow}, as

$$\sigma^2_{allow} = (0.3\sigma_{pt})^2$$

where σ_{pt} is the SD for PT assessment

6.2.2. Calculation of the critical value for the test as

$$c = F_1\sigma^2_{allow} + F_2 s^2_w$$

where
$F_1 = 1.88$ and $F_2 = 1.01$ (for 10 test samples measured in duplicate; 95% level of confidence). Refer to Table B.1 in ISO 13528:2022 for other values.

6.2.3. Use of the calculated between-sample standard deviation, s_s, in Step 6.1.2 and making decision based on the following criteria:

6.2.3.1. If $s_s > \sqrt{c}$

there is evidence at the 95% level of confidence that the between-sample standard deviation in the population of samples exceeds the allowable fraction of σ_{pt}; therefore, the test for homogeneity has failed.

6.2.3.2. If $s_s \leq \sqrt{c}$

there is no evidence at the 95% level of confidence that the between-sample standard deviation in the population of samples exceeds the allowable fraction of σ_{pt}; therefore, the test for homogeneity has been passed.

7. If the material does not fulfill the criteria for sufficient homogeneity:

7.1. If moisture analysis did not satisfy the criteria for homogeneity, reanalyze all parameters.

7.2. If 50% or more of the measurands failed the criteria for homogeneity, unpack, remix, repack, randomize, and reanalyze the sample and reevaluate the results.

7.3. If 50% or more of the measurands passed the criteria for homogeneity, repeat the conduct of analysis using the contingency samples.

7.3.1. If after analysis of contingency samples, 50% or more of the measurands failed the test for sufficient homogeneity, unpack, remix, repack, randomize, and reanalyze the sample and reevaluate the results.

7.3.2. If less than 50% of the measurands still failed the criteria for homogeneity, account for proficiency test item inhomogeneity and set a new σ'_{pt} using the following formula:

$$\sigma'_{pt} = \sqrt{\sigma^2_{pt} + s^2_s}$$

III. EVALUATION OF TEST MATERIAL STABILITY

The statistical method used in testing the stability of the proficiency test item is based on ISO 13528:2022 assessment criterion for a stability check.

The following are the steps in conducting the stability test:

1. **Selection** of at least three (3) test samples for each stability testing period in their final packaged form using systematic sampling,

2. **Preparation** of two (2) sub-samples from each test sample using techniques appropriate to the test material,

3. **Obtaining of a measurement result** on each of the sub-samples and completing the whole series of measurements under repeatability conditions (i.e. same laboratory, same analyst, same method and equipment),

4. **Testing for stability:**

   4.1 Computation of the general average of the measurements obtained in the homogeneity testing (i.e., 0 month). Designate the value as $\bar{y}_1$, .

   4.2 Computation of the general average of the measurements obtained in each stability testing period. Designate the value as $\bar{y}_2$.

   4.3 Use of the set σ_{pt} or computed σ'_{pt} and calculation of $0.3\sigma_{pt\ t}$ or $0.3\sigma'_{pt}$.

   4.4 Comparison of the absolute value of the difference between the general average of measurement in homogeneity testing, $\bar{y}_1$, and general average of the stability testing, $\bar{y}_2$, , with the critical value which is $0.3\sigma_{pt}$ or $0.3\sigma'_{pt}$.

      4.4.1 If $|\bar{y}_1 - \bar{y}_2| \leq 0.3\sigma_{pt}$ (or σ'_{pt}), the proficiency test item is stable

      4.4.2 If $|\bar{y}_1 - \bar{y}_2| > 0.3\sigma_{pt}$ (or σ'_{pt}), the proficiency test item is unstable

   4.5 If the material does not fulfill the criteria for stability:

      4.5.1 Expansion of the criterion for stability, to include the standard uncertainty of each measurement, using the following formula:

      If $|\bar{y}_1 - \bar{y}_2| \leq 0.3\sigma_{pt}$ (or σ'_{pt}) + $2\sqrt{u^2(\bar{y}_1) + u^2(\bar{y}_2)}$, the test item is stable

      where:
      σ_{pt}/ σ'_{pt} is the SD for PT assessment
      $u(\bar{y}_1)$ is the standard uncertainty of homogeneity test results
      $u(\bar{y}_2)$ is the standard uncertainty of stability test results

      4.5.2 If the proficiency test item is still unstable after expansion of criterion, issue *z'* score to account for instability.

REFERENCES

1. Analytical Methods Committee. "Robust statistics: a method of coping with outliers", AMC Technical Brief No. 6. 2001. Available online at <http://www.rsc.org/images/brief6_tcm18-25948.pdf> [April 30, 2007].

2. Hedges, S. and Shah, P. 2003. *Comparison of mode estimation methods and application in molecular clock analysis.* Available online at <http://www.biomedcentral.com/ content/pdf/1471-2105-4-31.pdf> [April 30, 2007]

3. Thompson, M., Ellison, S.L.R. and Wood, R. 2006. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories. *Pure Applied Chem.* **78** (1), 145-196.

4. International Organization of Standardization. *ISO/IEC 17043:2010 Conformity assessment – General Requirements for Proficiency Testing.* International Organization for Standardization (ISO), Geneva, Switzerland.

5. International Organization of Standardization. *ISO 13528:2022. Statistical methods for use in proficiency testing by Interlaboratory comparisons,* ISO, Geneva, Switzerland.

6. Food Analysis Performance Assessment Scheme (FAPAS). 2002. *Proficiency Testing Protocol: Organization and Analysis of Data*, 6th ed., FAPAS, York, UK.

7. *Random Error and Systematic Error.* Available online at <http://www.phys.selu.edu/rhett/plab193/labinfo/Error_Analysis/05_Random_vs_systematic.html> [April 30, 2007]

8. Burke, S. *Missing values, outliers, robust statistics & non-parametric methods.* Available online at <http://www.lgceurope.com/lgceurope/data/articlestandard/lgceurope/502001/4509/article.pdf> [April 30, 2007]

9. Hall, J.M., Bean, V.E. and Mattingly, G.E. 1999. *Preliminary Results from Interlaboratory Comparisons of Air Speed Measurements Between 0.3 m/s and 15 m/s.* 1999 NCSL Workshop & Symposium. Available online at <http://www.cstl.nist.gov/ div836/836.01/PDFs/1999/NCSL_056.pdf> [April 30, 2007].

10. Official Methods of Analysis of AOAC International. 2019.