INTERNATIONAL COOPERATION

ISO/IEC 17043 statistics ISO 13528- assigned value, homogeneity stability

April 2021

Homogeneity and stability ISO13528

- Establish criteria for **sufficient** homogeneity and stability. *To assure participants receive comparable test items*
- Homogeneity test have to be performent before the distrubution of test items
- Time valid for stability test -> till the end of the round
- If sufficient stability (or homogeneity) is not feasible need to consider an additional contribution to uncertainty

Homogeneity Check ISO13528

Selection of the **laboratory** that will perform tests and select the **method** as well

Select one or more parameters for the homogeneity assessment $Sr(test)/\sigma_{PT}$ less than 0,5 or $Sr/\delta_E < 1/6$

Preparation of sufficient **number** g (≥ 10) of test items for homogeneity tests (are selected among the test items of the round)

Homogeneity Check ISO13528

g items are divided into m ≥ 2 test portions

Testing of g x m test portions under repeatability conditions

Calculate mean, s_w (within sample standard deviation) και s_s (between samples standard deviation).

s_w<0,5 of σ_{pt} or 1/6 of δ_E .

What if replicate measurements are not feasible?

Homogeneity Check ISO13528

Other criteria

- Check measurement results in order of measurement to see for trends in the analysis

- Examine the results for PT items averages by production order- see for trends and decide for proper actions (differential PT items, discard the non homogeneous portions or other)

- Compare the S_s with σ_{PT} or δ_E: $s_s \leq 0{,}3\ \sigma p$ or $s_s \leq 0{,}1\delta E$

When σ_{PT} is not known in advance the Provider can check for differences between items (Analysis of variance at a=0,05) or use information from previous rounds, from precision experiments ISO5725-2) or ***accept the risk for inhomogeneity and check after the round when the robust σ_{PT} is known***

Harmonized Protocol for proficiency testing IUPAC, Pure and Applied Chemistry 78, 145–196, 2006

Testing for "sufficient homogeneity"

- When we test for so-called "sufficient homogeneity" in such materials, we are seeking to show that this variation in composition among the distributed units (characterized by the sampling standard deviation σ_{sam}) is negligible in relation to variation introduced by the measurements conducted by the participants in the proficiency test.
- As we expect the standard deviation of interlaboratory variation in proficiency tests to be approximated to by σ_p, the "standard deviation for proficiency assessment", it is natural to use that criterion as a reference value.
- The 1993 Harmonized Protocol required that the estimated sampling standard deviation s_{sam} should be less than 30 % of the target standard deviation σ_p,that is,

$$s_{sam} < \sigma_{all}$$

where the allowed sampling standard deviation $\sigma_{all} = 0.3\sigma_p$.

- At that limit, the standard deviation of the resultant *z*-scores would be inflated by the heterogeneity by somewhat less than 5 % relative, for example, from 2.0 to 2.1, which was deemed to be acceptable.
- If the condition were not fulfilled, the *z*-scores would reflect, to an unacceptable degree, variation in the material as well as variation in laboratory performance

Harmonized Protocol for proficiency testing IUPAC

To test for sufficient homogeneity, we have to estimate σsam from the results of a randomized replicated experiment by using ANOVA. In the experiment, each selected distribution unit is separately homogenized and analyzed in duplicate. Much depends on the quality of the analytical results.

If the analytical method is sufficiently precise, $\boldsymbol{\sigma_{sam}}$ can be reliably estimated, and any lack of sufficient homogeneity present detected with reasonably high probability. In fact, the test could be too sensitive.

The material can be significantly heterogeneous statistically, but the sampling variance negligible in relation to $\boldsymbol{\sigma_p}$. However, if the analytical standard deviation $\boldsymbol{\sigma_{an}}$ is not small, important sampling variation may be obscured by analytical variation.

We may get a nonsignificant result when testing for heterogeneity, not because it is not present, but because the test has no power to detect it.

So,

Recommendation 7

The analytical (repeatability) precision of the method used in the homogeneity test should satisfy $\sigma_{an}/\sigma_p < 0.5$ where σan is the repeatability standard deviation appropriate to the homogeneity test.

Trade off between between the cost of specifying very precise analytical methods and the risk of failing to detect important sampling variation

Harmonized Protocol for proficiency testing IUPAC-Homogeneity

The new statistical procedure

Rather than express the criterion for sufficient homogeneity in terms of the estimated sampling variance s^2_{sam}, as in the 1993 Harmonized Protocol, it is more logical to impose a limit on the true sampling variance σ^2_{sam}

It is this true sampling variance that is more relevant to the variability in the (untested) samples sent out to laboratories. Thus, the new criterion for sufficient homogeneity is that the sampling variance σ^2_{sam} must not exceed an allowable quantity $\sigma^2_{all} = 0.09\sigma^2_p$ (that is, $\sigma_{all} = 0.3\sigma_p$).

Then in testing for homogeneity it makes sense to test the hypothesis $\sigma^2_{sam} \leq \sigma^2_{all}$ against the alternative $\sigma^2_{sam} > \sigma^2_{all}$.

Outliers

Recommendation 9: Handling outliers in homogeneity testing in testing for sufficient homogeneity, duplicate results from a single distribution unit should be deleted before the analysis of variance if they are shown to be significantly different from each other by Cochran's test at the 99 % level of confidence or an equivalent test for extreme within-group variance.

Data sets containing discrepancies in two such distribution units should be discarded *in toto*. Pairs of results with outlying mean value but no evidence of extreme variance should not be discarded

Harmonized Protocol for proficiency testing IUPAC-stability

Materials distributed in proficiency tests must be sufficiently stable over the period in which the assigned value is to be valid. The term "sufficiently stable" implies that any changes that occur during the relevant period must be of inconsequential magnitude in relation to the interpretation of the results of a round.
Thus, if it were deemed that a change in the *z*-score of ± 1 would be inconsequential, then an instability amounting to a change in analyte concentration of $0.1\sigma_p$ could be tolerated.

Normally, the period in question is the interval between preparation of the material and the deadline for return of the results,although the period will be longer if residual material is to be re-used in subsequent rounds or for other purposes.

The stability test should involve exposure to the most extreme conditions likely to be encountered during the distribution and storage of the material, or to accelerated degradation conditions.

The material under test should be in the packaging in which it is to be distributed.
A comprehensive test for sufficient stability would be extremely demanding of resources
It is therefore not usually practicable to test every batch of material for every round in a series.

Harmonized Protocol for proficiency testing IUPAC-stability

It is a sensible prior precaution to test each new combination of material and analyte before it is first used in a proficiency test and occasionally thereafter.
It may additionally be useful to monitor stability by, for example, arranging for analysis of units pre and post-distribution by a single laboratory, providing for return of some distributed units for direct comparison with stored units, or comparing post-distribution analysis results with prior information such as homogeneity test data.

Basic stability tests involve a comparison of the apparent analyte levels between material subjected to likely decomposition conditions and material which has not been so treated. This usually requires a sample of the distribution units to be randomly divided into (at least) two equal subsets. The “experimental” subset is subjected to the appropriate treatment, while the “control” subset is kept under conditions of maximum stability, for example, at low temperatures and low oxygen tension.

Harmonized Protocol for proficiency testing IUPAC-stability

Alternatively, and especially if stability for extended periods is of interest, the control subset may be kept under ambient conditions while the experimental subset is kept under conditions of accelerated decomposition (e.g., higher temperatures). The materials are then analyzed simultaneously, or if that is impossible, as a randomized block design. Such experiments must be carefully designed to avoid compounding the effects of change in the material with variation in the efficacy of the analytical method used.

Analysis of the control material at the beginning of the test period and experimental material at the end automatically includes any run-to run analytical difference in the results and may well lead to the incorrect conclusion that there is a significant instability.

Harmonized Protocol for proficiency testing IUPAC-stability

The recommended approach is, if at all possible, to analyze the experimental and control subsets together, in a random order within a single run of analysis, that is, under repeatability conditions.

Any highly significant difference between the mean results of the two subsets can then safely be regarded as evidence of instability.

As in homogeneity testing, a conceptual distinction must be made between statistically significant instability and consequential instability. For instance, a highly significant change in the analytical results might be detected, but the change may still be so small that a negligible effect on the *z*-scores of the participants could be inferred. In practice, significance tests are not powerful enough to validate such a small instability unless an exceptionally precise analytical method is used, and/or inordinate numbers of distribution units analyzed. The stability test will, therefore, only detect a gross instability.

INTERNATIONAL
COOPERATION

Thank You Very Much For Your Attention.