amc recommendation

Editor: M Thompson. No 1 Jul 2004

Test for 'sufficient homogeneity' in a reference material

Scope

This recommendation covers the chemical composition of reference materials that are already divided into packaged units for distribution to different users. 'Sufficient homogeneity' means that the distribution units differ from each other in composition only to an extent that does not affect the interpretation of the results when the material is in its intended use.

Recommendation

1. The test shall be conducted on at least ten distribution units selected at random from all of the packaged units prepared for distribution.[a]
2. Each of the *m* selected units shall be homogenised, and duplicate test portions taken from it analysed, in a random order, by using a method of appropriate precision under repeatability conditions.[b,c]
3. The results of the analyses shall be scrutinised to ensure that anomalous features indicating problems with the analytical results (non-random features, lack of resolution) are absent. Datasets affected by such problems shall be discarded and the test repeated *ab initio*.[d, e]
4. Acceptable datasets shall be treated by one-way analysis of variance to provide estimates s^2_{sam}, s^2_{anal} of the between-unit variance and between-result variance respectively.
5. A criterion of compliance shall be related to an uncertainty u_f that is regarded as fit for purpose defined as an allowable variance $\sigma^2_{al} = 0.09u^2_f$. The material is deemed to be sufficiently homogeneous only if

$$s^2_{sam} \le \frac{\chi^2_{m-1}\sigma^2_{al}}{m-1} + \frac{(F_{m-1,m}-1)s^2_{an}}{2}$$

   where χ^2_{m-1} is the value exceeded with a probability of 0.05 by a chi-squared random variable with *m*-1 degrees of freedom, and $F_{m-1,m}$ is the value exceeded with a probability of 0.05 by a random variable with the F-distribution with *m*-1 and *m* degrees of freedom.[f]
6. When, among the differences between corresponding duplicate results, there is a single outlying difference, the respective pair of results may, after due consideration, be deleted from the dataset before the analysis of variance. No other data shall be deleted. If two or more such outlying differences are present, the whole dataset shall be discarded and the test repeated *ab initio*. [g]

Notes

a) 'Selected at random' means that a formal randomisation procedure shall be used rather than an informal procedure such as shuffling.

b) 'Appropriate precision' means that the repeatability standard deviation of results obtained by the method shall be smaller than 0.5 times the standard uncertainty u_f that describes fitness for purpose in the relevant application.

c) The balanced design recommended is probably optimally rugged. (In other words, there may be designs, requiring the same analytical outlay, that are slightly more powerful under ideal conditions of execution, but may be more prone to give misleading results under slight deviations of procedure.)

d) Non-random effects in the results can be mistaken for heterogeneity, or can mask it. Such effects result from a failure (i) to randomise the order of analysis or (ii) to use precision conditions other than repeatability (*e.g.*, when the analysis is split into several runs). If the analysis *cannot* be executed except in two or more runs, the analyst must make every effort to keep the conditions stable between runs.

e) Lack of digit resolution in the data may result in the variation between results or units being represented inadequately or not at all. Analysts conducting the tests may need to be advised to record more significant figures than customary.

f) A full account of the statistical rationale of the recommended method can be found in Reference 1.

g) Figure 1 below may assist in the recognition of the different types of outliers. Outlying differences are most easily detected by the use of Cochran's test at the 95% level of significance, although other statistically sound tests will suffice. A fuller discussion of outliers can be found below.

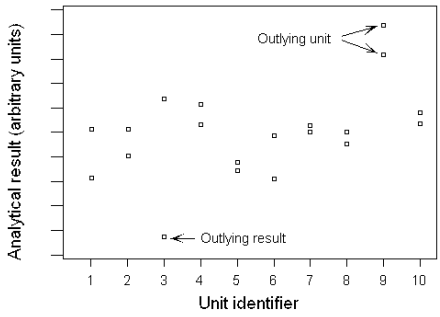

Fig. 1 Results from a test for sufficient homogeneity, showing two types of outlier.

Background to the Recommendation

Tests for homogeneity have traditionally been based on randomised repeated experiments followed by the use of analysis of variance to test the hypothesis $H: \sigma^2_{sam} = 0$. Unfortunately the outcome of such an experiment depends as much on the quality of the measurement as on the quality of the material. If the analytical precision in the experiment is low, the test may fail to detect heterogeneity. In contrast, a high-precision method may find as significant a degree of

heterogeneity that is entirely inconsequential. Moreover, most, perhaps all, candidate reference materials are actually or potentially heterogeneous and are demonstrably so, given a sufficiently precise analytical method. A naive test for heterogeneity *per se* is therefore pointless.

What is required is a test to demonstrate that the inevitable variation in composition between distribution units of the material is inconsequential in relation to it intended use. This requirement naturally calls for a comparison of the between-unit variance with an independent criterion based on fitness for purpose. An early test of this type [2] accepted the material as sufficiently homogeneous if $s_{sam} < 0.3u_f$, where, as above, s_{sam} is the estimated between-unit standard deviation and u_f is the standard uncertainty defining fitness for purpose. This test has the defect that the standard error of the estimated between-unit variance is highly dependent on the analytical precision. This factor could lead to an undue tendency of the test to reject material that was in fact satisfactory, a circumstance that adds considerably to the average cost of preparing a material.

The new test recommended here was designed to overcome that defect by taking account of the analytical precision. It tests the hypothesis $H : \sigma_{sam}^2 \le \sigma_{al}^2$, where σ_{al}^2 is allowable variance derived from the uncertainty regarded as fit for purpose. In many instances, $\sigma_{al}^2 = 0.09u_f^2$ is a suitable criterion.

Limitations of homogeneity tests
Tests for sufficient homogeneity can never be entirely satisfactory, however. The main problem is that any conceivable test will have a low statistical power (probability of rejecting the material when it is in fact unsuitable) in marginal cases unless a large $(m >> 10)$ number of units are analysed, at inordinate expense. It is therefore better to keep the cost of the test within reasonable bounds by minimising the false rejection of satisfactory material and regarding the test as a filter to prevent gross mistakes.

Homogeneity of individual units
This Recommendation does not advocate experimental designs that test the contents of individual units for homogeneity. Although such a procedure is currently recommended by ISO [3], it doubles the cost of analysis (for balanced designs) and the outcome is of no practical consequence. A unit that is 'homogeneous' at the time of testing cannot be guaranteed to be so after transportation, handling and storage. It is therefore a universally recognised responsibility of the individual analyst to ensure that any laboratory sample, including the special case of the reference material, is sufficiently homogeneous before the test portion is taken.

Types of outliers
There are potentially two types of discrepant result stemming from the test for sufficient homogeneity, namely outlying units and outlying results, as shown in Figure 1. An outlying unit is indicated by two concordant results from the unit, the mean of which is discordant with the means of the remaining pairs of results. **A pair of outlying results from a single unit is indicative of heterogeneity and the results must not be discarded before analysis of variance is applied.** Outlying results occur when the two results stemming from a single unit differ from each other by an amount that is implausible in relation to all of the other differences between pairs of results. **Contrary to previous recommendations** [2], **it is usually appropriate to delete a pair of results showing an outlying difference and then proceed with the analysis of variance.**

Rationale for deleting an outlying result
The justification for deleting an outlying pair of results is as follows. Analytical outliers resulting from uncontrolled variation in the analytical procedure are well recognised and occur at a rate of perhaps a few percent. A dataset from a homogeneity test contains at least 20 results, so datasets containing a single outlier (but otherwise reliable) are encountered quite often. Analytical outliers have the effect (perhaps unexpected *prima facie*) of making the material more likely to pass the test for sufficient homogeneity, so cannot be tolerated. However, rejecting the whole dataset because of a single outlier would be unduly costly.

The situation can be saved by the following consideration. Analytical good practice requires that the test material of each unit is homogenised to an appropriate degree before the test portions are taken. Therefore discrepant duplicate results from a properly homogenised single unit must result from analytical operations performed after the test portion is taken and do not reflect on the condition of the material under test.

It is therefore usually sensible to discard from the dataset one pair of such results shown to be significantly different (for example by Cochran's test) before applying analysis of variance. The occurrence of two such pairs of results in a dataset, however, calls into question the reliability of the whole analytical operation: the dataset should be discarded and the whole test repeated *ab initio*.

Due consideration must be given to the nature of the material before an outlying result is discarded. There are rare but well known instances (*e.g.*, ores of precious metals) where the analyte is present at low trace levels overall in the bulk material but in the form of just a few particles containing the analyte at a concentration approaching 100 %. In such a case, the presence of an outlying result might result from this extreme type of heterogeneity rather than an analytical outlier. More elaborate investigations than the current Recommendation are required for this type of material.

References

1. T Fearn and M Thompson, *Analyst*, 2001, **126**, 1414-1417.
2. M Thompson and R Wood, *Pure Appl Chem*, 1993, **65**, 2123-2144.
3. ISO Standard on preparation of reference materials, Part No???

This recommendation was prepared for the AMC by the Statistical Subcommittee.