Sampling theory and sampling uncertainty

Analytical Methods Committee, AMCTB No. 71

We make a chemical measurement mostl to help make a rational decision about a 'target', a particular mass of material that is of interest in manufacturing, commerce, human health, or for cultural purposes. A target might comprise for e ample a shipment of a ra material,

experimentally before the sampling protocol could legitimately be accepted as appropriate. A further problem needs to be considered. Successive targets, especially of unprocessed materials, differ from each other in numerous ways, so a protocol that delivers a suitable sample from one target may do otherwise for the next one ostensibly of the same kind.

It is difficult, however, to fault the TS as a qualitative method of arriving at what is prima facie a reasonable procedure, except perhaps on the grounds of the effort required. Much of the theory is commonsensical and, moreover, the process will be educational for trainee samplers. However, the sampling procedure thus arrived at will need validation (and possibly some amendment) before it can be accepted as fit for purpose. This is because it is difficult indeed and often very laborious to quantify many of the 'errors' (not to mention their interactions, which are usually ignored), so the 'correctness' cannot be taken for granted. Furthermore the aim of TS is less to make an explicit estimate of the uncertainty arising from the sampling than to provide a 'representative' sample that can be sent to a laboratory without contributing any apparent uncertainty.

The e perimental school of thought

The alternative school of thought holds that, in a properly randomised experiment, simply replicating the application of any sampling protocol gives a useful estimate of the uncertainty of the resultant measurements arising from sampling. (That is why the strategy is sometimes confusingly called the 'Measurement Uncertainty' (MU) approach.) The protocol under test could be arrived at by any means: by tradition, by an evolutionary process, from TS, or simply by judgement based on experience. If properly conducted, the replication can encompass much of the potential uncertainty and lets us judge whether the protocol is fit for purpose. (The designs shown below, however, cannot incorporate uncertainty relating to operator/method bias.)

A parsimonious experimental approach is to make randomised duplication a part of routine sampling (by using a provisional protocol) until the required amount of data is obtained. This ensures that the uncertainty estimate obtained represents real-life conditions rather than an artificial experimental situation. The design shown in Fig. 1 (or an even more economical unbalanced version) is appropriate. Results are collected until there are enough to allow a reasonably stable estimate of the between-sample variance by hierarchical ANOVA (analysis of variance). (After that, the occasional duplicate sampling of a target can be regarded as merging into internal quality control of sampling.) A set of results from such a test might resemble those depicted in Fig. 2.

A careful visual examination of the data is an essential preliminary step, to ensure that a suitable statistical approach is employed. In Fig. 2 we see successive targets of similar composition apart from one possibly anomalous target (no. 6). However, a single anomalous target per se will not affect the nested ANOVA because the between-target dispersion is not relevant here. Between-sample variation is apparently greater than analytical variation. There is no suggestion of

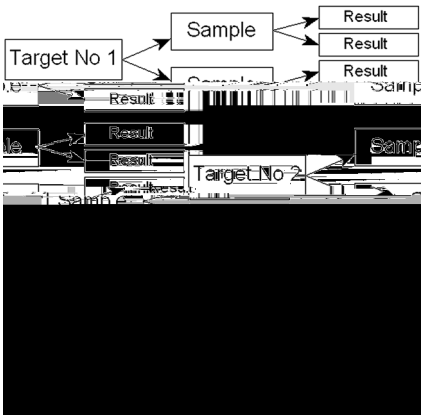

Fig. 1 Design of a balanced duplicated sampling experiment. An unbalanced design reduces the analytical burden by 25% (see AMCTB no. 64).

heteroscedasticity or that the first sample differs systemically from the second. Target no. 7 has the biggest difference between samples but it is not clear visually that the difference is outlying. Either way, a robust ANOVA can cope with this dataset, providing an estimate for the 'typical' value of the between-sample standard deviation. The statistics obtained were: grand mean, 11.1% mass fraction, between-target SD, 0.15; within-target/between-sample SD, 1.01; analytical (within-sample) SD 0.32.

In instances where the results are heteroscedastic (that is, the analytical and/or sampling standard deviation varies with the concentration of the analyte) a more complex type of statistical analysis may be required. Fig. 3 shows such a dataset. It is evident there that the dispersion of both analytical and sample duplicates is greater at high than at low concentrations. A suitable treatment for this particular dataset might be log-transformation before ANOVA is attempted. That would tend to stabilise the variance, a requirement for a usable outcome of ANOVA. An examination of the residuals would show whether that strategy had been successful.

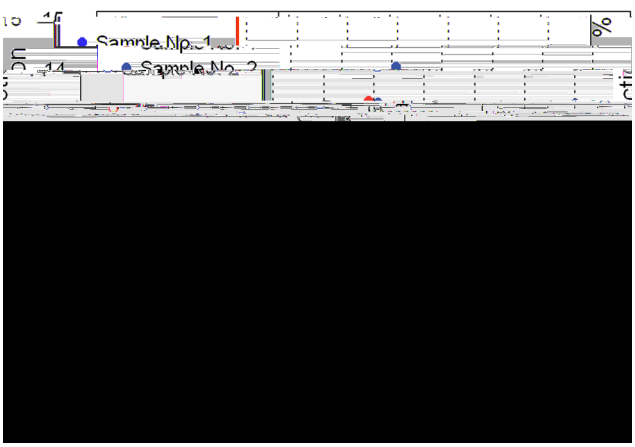

Fig. 2 Results from a duplicated multi-target experiment.

Limitations of the e perimental approach

A clear shortcoming of replication is that, in the event of the protocol being rejected as being unfit for purpose, we have no immediate diagnostic information to locate and rectify the source of the problem. Further experiments would be required. In addition, we have already seen that the duplicate method fails to incorporate sampler bias and method bias, and for the present time we have perforce to accept that circumstance. It is fi