amc technical brief

Analytical Methods Committee No. 16 Apr 2004 © Royal Society of Chemistry 2004

Proficiency testing: assessing z-scores in the longer term

While a single z-score provides a valuable indication of the performance of a laboratory, a set or sequence of z-scores provides a deeper insight. Moreover, z-scores collected over time for a single analyte/test material combination provide information on the participant's uncertainty. Both graphical and numerical methods may be appropriate for assessing a sequence of z-scores. However, due caution is required with numerical methods to avoid incorrect conclusions. The use of a summary score derived from z-scores relating to a number of different analytes is not recommended: it has a very limited range of valid applications and tends to conceal sporadic or persistent problems with individual analytes. Moreover it is prone to misuse by non-scientists.

Summary scores

The following two types of summary score are statistically soundly based [1] and may be useful for individual participants to assess a sequence of z-scores $z_i, i = 1,\ldots,n$ derived from a single combination of analyte, test material and method.

The rescaled sum of the z-scores,

$$\mathrm{RSZ} = \sum_i z_i \Big/ \sqrt{n} \,,$$

can be interpreted on the same basis as a single z-score, *i.e.*, it is expected to be zero-centred with unit variance if the z-scores are. This statistic has the useful property of demonstrating a persistent bias or trend, so that the sequence of results [1.5, 1.5, 1.5, 1.5] provides a statistically significant RSZ of 3.0, even though any single one result is not significant at the 95% confidence level. However, RSZ could conceal two large z-scores of opposite sign that roughly cancel.

The sum of the squared z-scores,

$$\mathrm{SSZ} = \sum_i z_i^2 \,,$$

could be interpreted as a χ_n^2 distribution for zero-centred z-scores with unit variance. This statistic has the advantage of avoiding the cancellation of large z-scores of opposite sign, but is less sensitive to small biases. Both of these summary statistics need to be protected (for example, by robustification or filtering) against past outlying scores, which would otherwise have a long-term persistence. SSZ is especially sensitive to outliers.

Uncertainty

Both of these statistics (when so robustified) can be related to uncertainty of measurement in the following way. If the z-scores are based on fitness for purpose and therefore ideally $N(0,1)$ for compliant laboratories, significantly high levels of the summary statistics indicate that the participant's uncertainty of measurement is greater than indicated by the schemes fitness-for-purpose criterion.

Control charts

Graphical methods of summarising a set of z-scores can be just as informative and probably less prone to misinterpretation than summary scores. Multiple univariate charts [2], such as those shown overleaf, give a clear overview and are especially useful when scores from a group of analytes determined by a common method are considered. Hand-drawn charts are quick to update and serve just as well as those produced by computer.

The control chart (Figure 1 overleaf) shows upward-pointing symbols to indicate z-scores greater than zero and downward-pointing symbols for those less than zero. Small symbols represent instances where $2 \le |z| < 3$, and large symbols instances where $|z| \ge 3$. The data illustrated immediately show some noteworthy features. Results from round 11 are mostly too low, demonstrating a procedure that was faulty in some general feature, while analyte 7 gives high results too frequently, demonstrating a persistent problem with that specific analyte. The remaining results are roughly consistent with fitness for purpose, which on average would result in about 5% of z-scores represented by a small symbol.

J-charts

A J-chart (otherwise known as a 'zone chart') [3] is even more informative, because it combines the capabilities of the Shewhart and the cusum charts. It does this by cumulating special J-scores attributed to successive results on either side of the zero line. This enables persistent minor biases to be detected as well as abrupt large changes in the analytical system. Typical rules for converting z-scores to J and cumulating them are as follows.

- If $z \ge 3$ then J = 8.
  If $2 \le z < 3$ then J = 4.
  If $1 \le z < 2$ then J = 2.
  If $-1 < z < 1$ then J = 0.
  If $-2 < z \le -1$ then J = –2.
  If $-3 < z \le -2$ then J = –4.
  If $z \le -3$ then J = –8.
- J-scores from successive rounds are cumulated until $|z| \ge 8$, which defines an excursion beyond the action limits, and triggers investigative procedures.
- The cumulator is reset to zero after any such excursion, and when successive values of J are of opposite sign.

Several examples of the cumulative effect of bias are visible in Figure 2 (which illustrates the same results as Figure 1 for comparison). For example, Analyte 3 in Rounds 1-4 receives z-scores of 1.5, 1.2, 1.5, and 1.1 respectively, translating into J-values of 2, 2, 2, and 2, which cumulate to 8 by Round 4 and trigger investigative procedures.

References

1. M Thompson and R Wood, *Pure Appl Chem*, 1993, **65**, 2123-2144.
2. M Thompson, K M Malik and R J Howarth, *Anal Comm*, 1998, **35**, 205-208.
3. *AMC Technical Briefs*, No 12 (2003), www.rsc.org/lap/rsccom/amc/amc_index.htm

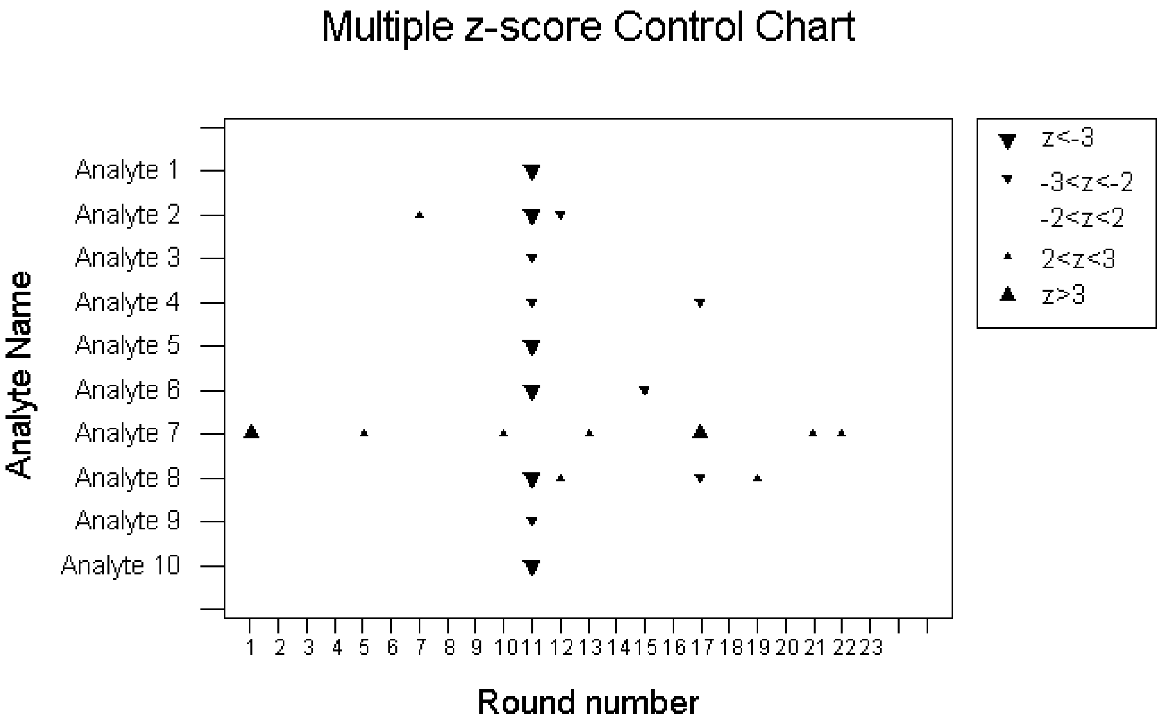

Fig 1. A multiple control chart for z-scores.

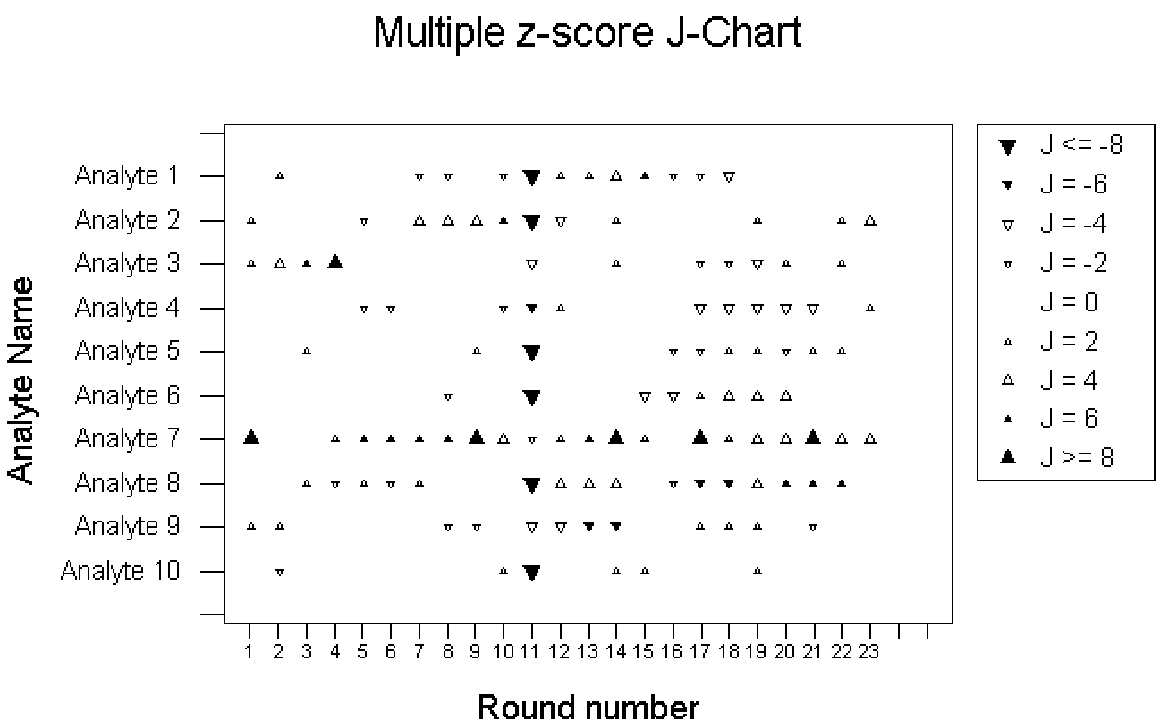

Fig 2. A multiple J-chart for z-scores (same data as Figure 1).

Other AMC Technical Briefs can be found on: **www.rsc.org/lap/rsccom/amc/amc_index.htm**