amc technical brief

Analytical Methods Committee No.12 Mar 2003

The J-chart: a simple plot that combines the capabilities of Shewhart and cusum charts, for use in analytical quality control

Internal quality control (IQC), based on the interpretation of results from the analysis of one or more control materials in every analytical run, is an essential part of routine analysis. The interpretation of the results is usually based on the well-known Shewhart and cusum charts.[1] The J-chart (or zone control chart)[2, 3] is a very useful but little known alternative. It combines the information provided by both of the traditional charts, very effectively detecting both abrupt changes and drift in the result obtained. It is easy to set up and apply. Although computer-based implementations exist, the J-chart is equally well suited to manual charting methods.

The interpretation of IQC results, a universal concern in routine analytical operations, ensures that the ongoing results continue to comply with fitness-for-purpose goals and that abrupt departures from expected values or an onset of drift can be promptly detected. The 'zone control chart'[2,3] (also called the J-chart[4]) was devised as a simple and very effective tool for monitoring routine analytical IQC operations, where results obtained on one or more control materials are collected in each run of analysis. The J-chart combines aspects of both the traditional Shewhart control-chart for the mean, and cumulative-sum control schemes. (See reference 4 for a review of univariate quality-control charting methods.) Although the J-chart is implemented in some statistical packages, *e.g.*, SAS® and MINITAB® (where it is called the zone chart), the method is also simple to carry out by hand.

Establishing the reference values

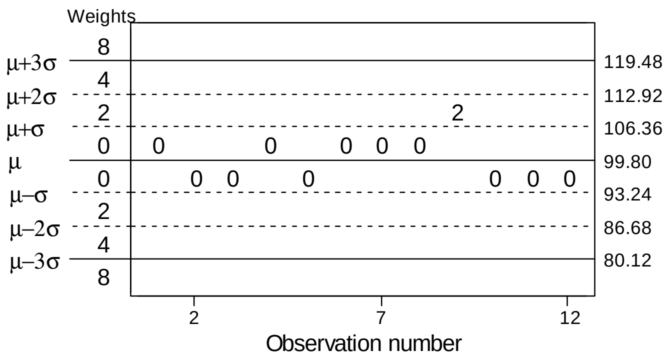
In IQC, as with any quality control method, some preliminary work must be undertaken to establish reliable estimates of the parameters describing the control material. Those are the mean result μ and the standard deviation σ representing run-to run variation. (In other words, the replicated results should be obtained from separate runs of the analytical procedure - repeatability standard deviation would be too small.) Moreover, the results must be obtained when the analytical system is operating under statistical control.

The estimates should therefore be based on at least 10 prior observations (x_i). The observations can be either single results or within-run means (based on a fixed number of replicates) that may be considered as individual 'observations' for IQC purposes. Care should be taken to remove or downweight any obvious outliers from the data obtained during this 'training' stage, before the mean and run-to-run standard deviation are calculated. An additional measure that can be used to obtain a better estimate is to implement Nelson's suggestion[5], which minimises inflationary effects on variability caused by any local trends or oscillations present within the data. In Nelson's method, the estimate of σ is obtained by calculating the average moving range ($\overline{MR}$), *i.e.*, taking the sum of successive pair differences (x_2 - x_1), (x_3 - x_2), *etc*. without regard to sign, and dividing it by the total number of pairs. The standard deviation is then given by $\sigma = 0.8865 \times \overline{MR}$.

The J-chart

Having defined both μ and σ, we are now in a position to set up the J-chart. Its vertical axis is in units of σ and its horizontal axis is time. The horizontal centre-line corresponds to μ. Three equal-width bands, or 'zones'[2,3], are marked-off horizontally on each side of the centre-line; their boundaries correspond to the values of $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$. These values can conveniently be written, for reference, at the ends of the boundaries where they meet the vertical axis of the chart (Figure 1).

Figure 1



Suppose the first observation from the newly operational analytical system is now made (x_1).

- If the observation x_1 falls between $\mu - \sigma$ and $\mu + \sigma$ it is assigned a score of zero.
- If x_1 falls between $\mu + \sigma$ and $\mu + 2\sigma$, or between $\mu - 2\sigma$ and $\mu - \sigma$, it is assigned a score of 2.
- If x_1 falls between $\mu + 2\sigma$ and $\mu + 3\sigma$, or between $\mu - 3\sigma$ and $\mu - 2\sigma$, it is assigned a score of 4.
- If x_1 falls in either of the outermost zones (greater than $\mu + 3\sigma$ or less than $\mu - 3\sigma$) it is assigned a score of 8.

As successive observations become available, their scores are cumulated, and the cumulative sum is written in the centre of the zone at the appropriate point on the chart. However, as soon as a new observation falls on the *opposite* side of the centre-line to the immediately preceding observation, the total score is reset to zero and cumulation then starts again, beginning with the current score. Once the total score equals or exceeds 8, the system is deemed to be 'out-of-control', and the cause of the problem should be investigated before resumption of analysis.

Although alternative scoring schemes have been suggested (and can be used, for example, in the Minitab implementation), the simplicity of Jaehn's scheme has been found both to appeal to the user and to perform satisfactorily in practice. In computer-based implementations of the scheme, which do not necessarily require graphical display, it is informative to designate scores above the centre-line as positive and those below as negative, to distinguish out-of-control situations of each type.

Example 1
The training set consists of 20 successive computer-generated values drawn at random from a normal distribution with mean 100.0 and standard deviation 5.0, rounded to 1 decimal place. Reading from left to right we have:

105.9	102.9	94.7	99.9	100.3
99.4	99.7	90.7	113.3	95.9
101.9	102.7	98.7	103.3	96.6
96.7	107.3	93.2	86.5	106.5.

Here μ = 99.80, $\overline{MR}$ = 7.40 and σ = 6.56.

The test set consists of the next 12 simulated values. Reading across, we have:

104.6	99.0	98.4	103.5	93.5
104.0	102.0	100.9	112.3	96.1
99.0	96.9.			

The corresponding J-chart is shown in Figure 1 (above). As would be expected for random normal data, the simulated 'analytical system' appears to be in complete control.

Example 2
This example uses results for ^{27}Al (ppb) taken from a multi-run dataset for a NIST foodstuff reference material. A preliminary study of the total multi-isotope data showed that a few of the early determinations, which necessarily form the training-set, were multivariate outliers; the observations corresponding to those 5 sets of multi-isotopic determinations were therefore deleted from the training-set. The final training-set values for ^{27}Al (reading across) were:

245253	221548	227207	213298	228872
212280	223115	185191	207478	212904
186244	219228	202954	221978	224347
200476.				

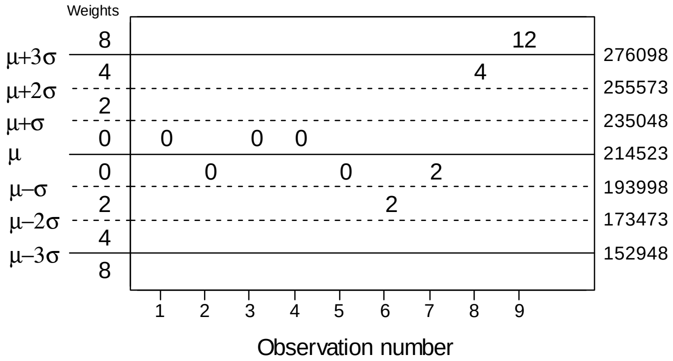
Here μ = 214523, $\overline{MR}$ = 23153 and σ = 20525 ppb.

The results of the first 9 determinations in the ensuing test-set were (again reading across):

219228	202954	221978	224374	200476
192291	199859	263992	276790.	

The corresponding J-chart is shown in Figure 2.

Figure 2



It appears that following observation 6 the measurements have begun a systematic drift. This reaches 'out-of-control' magnitude (with a cumulative score of 12) by observation 9. In such an instance, the cause of the apparent drift should be investigated and rectified before further analysis is undertaken, in accordance with standard practice in IQC.

References

1. M Thompson and R Wood, *Pure Appl. Chem.*, 1995, **67**, 649-666.
2. A H Jaehn, *Technical Association of the Pulp and Paper Industry Journal*, 1987, **87**, 159-161.
3. A H Jaehn, *Quality Progress*, 1991, **24**, 65-68.
4. R J Howarth, A*nalyst*, 1995, **120**, 1851-1873.
5. L S Nelson, *Journal of Quality Technology*, 1982, **14**, 172-173.

This Brief was prepared for the AMC by Richard J Howarth.