Advice Sheet 37: Understanding and Interpreting PT Scheme Scores

Introduction

PT schemes involve the participant laboratories receiving, on a regular basis, a sample with unknown concentrations of the analytes of interest, analysing the samples and returning their results to the scheme organiser. The scheme organiser prepares a report which includes a statistical analysis of how the laboratory's results compare to those of the other participants. This information, used correctly, can provide a powerful tool for the laboratory to monitor the quality of the results it produces and give their clients confidence that the data they receive are fit for purpose. However, care must be exercised when interpreting the statistical aspects of the report (usually presented as z-scores) as there are many factors which can compromise the results or even render them useless. This is exacerbated by the fact that different PT scheme providers use different statistical methods and apply different rules in interpretation of the data provided by the laboratories.

The z-score

The z-score is a ratio presented as a single number which, at its simplest, represents the number of standard deviations the laboratory's result is away from the 'true' value assigned by the scheme provider.

$$Z = \frac{(laboratory\ result - assigned\ value)}{standard\ deviation}$$

A negative z-score indicates the laboratory's result was below the assigned value and a positive z-score indicates that the laboratory's result was above the assigned value. The scheme provider can then set cut-off points for the numerical value of z which place the laboratory's data into bands such as 'satisfactory', 'questionable', 'unsatisfactory'.[1]

What should be noted is that there is no one way to do this: different PT scheme providers use different statistical models (the data sets returned are rarely simple) and may band the z-scores differently from one another.

Typically a laboratory would investigate an unsatisfactory result and might investigate recurring questionable results.

However, in order for the interpretation of a z-score to be meaningful the components of the calculation (assigned value, laboratory result and standard deviation) need to be fit-for-purpose. Assuming the laboratory has supplied a valid result (correct units, the right determinand etc.) then the validity of the assigned value and the standard deviation (both of which are usually derived from a statistical analysis of the results supplied by all participant laboratories) can be considered as follows.

Validity

The first statistical parameter to consider is the simplest one: the number of data points. A low number of data points make meaningful statistical analysis impossible and for this reason most PT scheme

[1] These are the most common ones.

t: +44(0)1344 886338 w: www.cawood.co.uk/nrm
NRM, Coopers Bridge, Braziers Lane, Bracknell RG42 6NS

CDA-037-TBA-V2
Issued: 11th August 2021

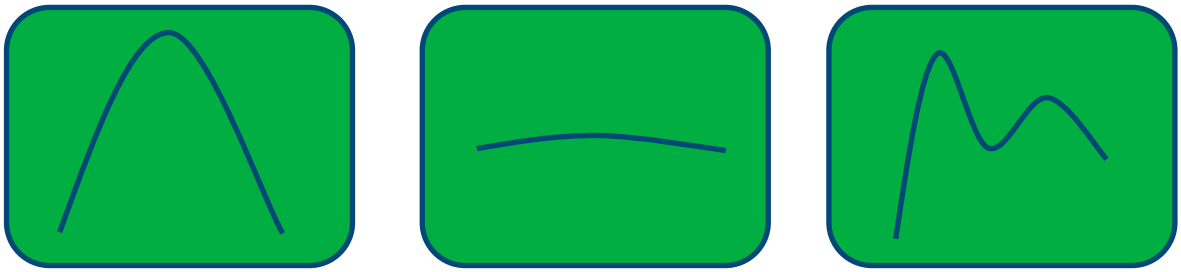
providers do not supply a z-score and/or banding if there are only a few points available. Schemes generally cut off at around eight points or fewer, either not providing a z-score or removing the banding.

Scheme providers employ complex statistical analysis to determine the assigned value and the standard deviation. One exception is where the test sample is wholly artificial, for instance a pure solution the composition of which is known very precisely: in this case the assigned value is the concentration of the solution. Of more value are real world samples (in our case soils, sludges, food products for example): these are much closer to the actual samples but the actual concentration cannot be known in advance and so the scheme provider must rely on the data provided by the participant laboratories.

In cases where the assigned value cannot be known in advance (which is most of the time) the scheme provider uses statistics to assign a 'correct' value: at its simplest this is often the median – or some measure of the mid-point of the data supplied by the laboratories (the *arithmetic* mean is rarely used as it is prone to influence from outliers in the data). The standard deviation (a measure of the spread of the data) is then also determined by the scheme provider. This is often an assessment of what is fit-for-purpose.

If the standard deviation is too large then there is most likely no clear assigned value as the points obtained from the laboratories are too spread out. Another scenario which can occur is where more than one assigned value is possible. This often happens when the laboratories use multiple methods to provide the same result; for example total copper might be determined by ICP or AAS. Different methodologies have different inherent biases and this may lead to the behaviour shown in Figure 1(c) below. Again in cases (b) and (c) the PT scheme provider may remove the z-score and/or banding from their report.

Figure 1: Some common distributions. The x-axis is the analytical result, and the y-axis is the number of laboratories achieving that analytical result.



(a) (b) (c)

Figure 1 highlights the importance of a visual inspection of the data. All of these distributions can produce a z-score for each laboratory but only (a) has a clearly defined value which can be used as an assigned value. In (b) the data is too spread out and in (c) there are two possible assigned values.

Most schemes now provide some graphical representation of the data in addition to just the raw z-score. It is important therefore that a customer take into account the complex nature of PT scheme data when assessing z-scores provided by a laboratory. This is best done by working with the laboratory to investigate the data underlying any z-score, particularly those which may indicate questionable or unsatisfactory performance.

t: +44(0)1344 886338 w: www.cawood.co.uk/nrm
NRM, Coopers Bridge, Braziers Lane, Bracknell RG42 6NS

CDA-037-TBA-V2
Issued: 11th August 2021