Applications of robust covariance estimators in examination of inter-laboratory study data

Stephen L R Ellison

LGC Limited, Queens Road, Teddington, Middlesex TW11 0LY

Introduction

Many inter-laboratory studies collect results for more than one measurand from each participant:

- **PT rounds** can use multiple samples, multiple analytes, or both
- **RM certification** can include many analytes
- **Validation studies** involve multiple test materials

Inspecting multivariate data can be hard. Single outliers are easy to find (Fig 1) but some anomalies (such as interchanged, similar, samples or a consistent problem for one lab) only show as a pattern.

Multivariate outlier detection methods can simplify inspection. Most multivariate methods require covariance information, but traditional covariance estimates are badly affected by outliers.

This poster illustrates the use of some robust estimates of covariance to improve outlier detection in multivariate inter-laboratory data.

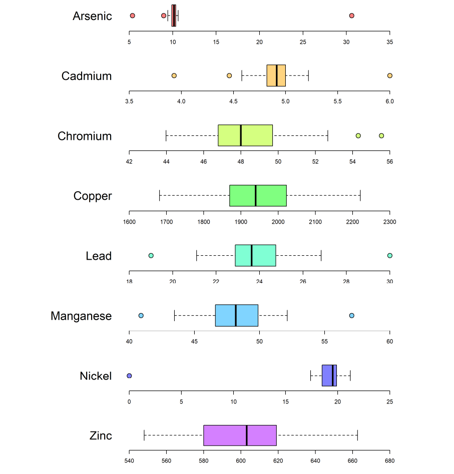
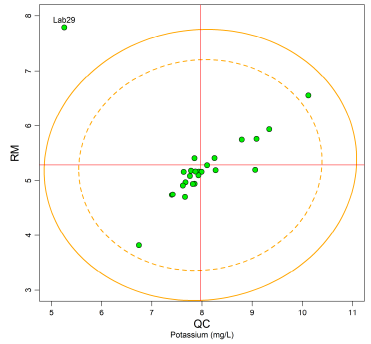


Figure 1: Lab medians (5 replicates) for eight elements in a reference material study

Improved Youden plots

Youden plots [3] show results for one sample (or analyte) plotted against another. Figure 2 shows that a robust covariance estimate sharpens outlier detection considerably.

a)



b)

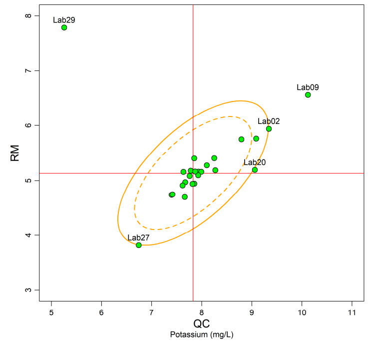


Figure 2: Confidence ellipses for paired data based on a) Pearson correlation and b) the GK covariance estimator. The traditional estimator substantially over-estimates both variance and covariance due to outlying values.
Laboratory 29 shows (also for other elements) evidence of accidental interchange of the two test materials.

Robust Mahalanobis distance

Mahalanobis distance (MHD) is a measure of distance from a (multivariate) centre, scaled by a complete covariance matrix. Figure 3 shows Mahalanobis distances for the laboratory data of Figure 1.

The conventional covariance (Fig 3a) is severely inflated by outliers, giving low MHD values that fail to distinguish important anomalies.

Basing the distance on a robust covariance matrix (Fig 3b) greatly improves anomaly identification.

a) Conventional b) Robust (MCD)

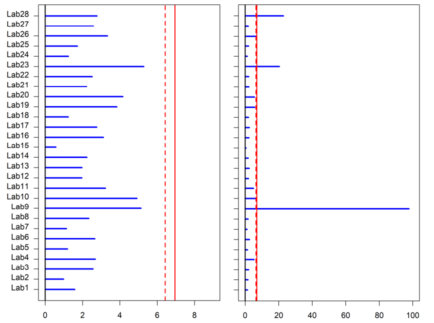


Figure 3: Mahalanobis distance *D* for the data of Figure 1. a) conventional covariance; b) MCD covariance estimate. Red lines are 95% (dashed) and 99% (solid) upper limits based on the χ-squared distribution.
Note that distances and limits are shown as distance and not distance squared

Robust estimators used in this poster

Two robust covariance estimators are used here for illustration, though many more exist:

GK [1] A simple estimator that can use any robust standard deviation estimator s^*:

$$cov_{GK}(x,y) = [s^*(x+y)^2 - s^*(x-y)^2]/4$$

MCD The minimum covariance determinant method introduced by Rousseew et al. [2]

Conclusions

The availability of robust covariance estimates allows the implementation of effective new methods for inspection of multivariate inter-laboratory data.

Improved tools include enhanced Youden plots and robust variants of Mahalanobis distance.

These tools permit identification of anomalies that might go undetected by univariate methods.

References

1. Gnanadesikan, R., Kettenring J. R. (1972) Biometrics 28, 81-124
2. Rousseeuw P. J. , van Driessen K. (1999) Technometrics 41, 212-223
3. Ellison SLR. (2018). metRology: Support for Metrological Applications. R package version 0.9-27. http://metrology.sourceforge.net