TOLERANCE ZONES THROUGH ROBUST CLUSTERING TECHNIQUES

L.A. García-Escudero and A. Gordaliza
Departamento de Estadística e Investigación Operativa
Universidad de Valladolid. Spain

A technique for obtaining (asymptotic) distribution free tolerance intervals was introduced in Butler (1982). Those intervals turn to be the ones containing the non-discarded observation when computing the well-known Rousseeuw's Least Trimmed Squares (LTS) estimators in the univariate location case. Butler, Davies and Jhun (1993) also propose a tolerance ellipsoid based on the Minimum Covariance Determinant (MCD) multivariate location and scatter estimator.

In a similar fashion, tolerance zones can be obtained from the non-trimmed observations after a trimmed k-means procedure. Trimmed k-means were introduced in Cuesta-Albertos, Gordaliza and Matrán (1997) with the aim of robustifying the classical k-means clustering procedure.

In this work, we explore how these (non necessarily connected) tolerance regions can be applied in the Quality Control framework. As the procedure works out nicely for moderately high dimensional (non-standard) situations, this approach could be useful in some engineering applications where this sort of data appears.

References

[1] Cuesta-Albertos, J.A., Gordaliza, A. and Matrán, C. (1997), "Trimmed k-means: An attempt to robustify quantizers," *The Annals of Statistics,* **25**, 553-576.

[2] Butler, R.W. (1982), Nonparametric tolerance interval and point prediction using data trimmed by a Grubbs-type outlier rule. *Ann. Statist.* **10**, 197-204.

[3] Butler, R.W., Davies, P.L., and Jhun, M. (1993). Asymptotics for the minimum covarince determinant estimator. *Ann. Statist.* **21** 1385-1400.