

Amendment to the Classification Macro (July 2019)

The macro named 'classification' previously provided summary results for the cubic clustering criterion or CCC statistic (Sarle, 1983) for indicating the appropriate number of clusters. However, this macro now also includes the same summaries for the Pseudo-F statistic (Calinski and Harabasz, 1974). As recommended in Section 6.2 of the users' guide, a positive CCC statistic > 2 will still be necessary to ensure that clusters are well-defined and in most cases, the maximum values for the two statistics will coincide and suggest the same number of clusters. However, Milligan and Cooper (1985) have shown that the Pseudo-F is somewhat better in determining the correct number of clusters. Therefore, when the maximum values of the CCC and Pseudo-F statistics suggest different numbers of clusters, the number suggested by the Pseudo-F is recommended. Also, note that neither statistic may perform well for clusters that are very elongated.

Sarle, W.S. *Cubic Clustering Criterion* (SAS Technical Report A-108). Cary, NC: SAS Institute; 1983.

Calinski, T. and Harabasz, J. A dendrite method for cluster analysis. *Communications in Statistics* 1974; 3:1-27.

Milligan, G.W. and Cooper, M.C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 1985; 50:159-179.