

index.G1(clusterSim)

### **Caliński-Harabasz pseudo F-Statistic**

(for metric data only)

$$G1(u) = \frac{\text{trace}(\mathbf{B})/(u-1)}{\text{trace}(\mathbf{W})/(n-u)},$$

where:  $\mathbf{B}$  – between-group dispersion matrix,

$\mathbf{W}$  – within-group dispersion matrix,

$u$  – number of clusters ( $u = 2, \dots, n-1$ ),

$n$  – number of objects.

The value of  $u$ , which maximizes  $G1(u)$ , is regarded as specifying the number of clusters.

### **References**

Caliński, R.B., Harabasz, J. (1974), *A dendrite method for cluster analysis*, „Communications in Statistics”, vol. 3, 1-27.

Everitt, B.S., Landau, E., Leese, M. (2001), *Cluster analysis*, Arnold, London, p. 103.

Gatnar, E., Walesiak, M. (Eds.) (2004), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych [Multivariate statistical analysis methods in marketing research]*, Wydawnictwo AE, Wrocław, p. 338.

Gordon, A.D. (1999), *Classification*, Chapman & Hall/CRC, London, p. 62.

Milligan, G.W., Cooper, M.C. (1985), *An examination of procedures of determining the number of cluster in a data set*, “Psychometrika”, vol. 50, no. 2, pp. 159-179.