dist.SM(clusterSim)

Sokal-Michener distance measure d_{ik} for nominal data

$$d_{ik} = \frac{\sum_{j=1}^{m}(1 - g_{ik}^{(j)})}{m} = \frac{m - m_r}{m},$$

where: $i, k = 1, \ldots, n$ – the number of object,
$j = 1, \ldots, m$ – the number of variable,
x_{ij} (x_{kj}) – *i*-th (*k*-th) observation on *j*-th variable,
$g_{ik}^{(j)} = \begin{cases} 1, & x_{ij} = x_{kj} \\ 0, & x_{ij} \neq x_{kj} \end{cases}$,
m_r – number of variables where $x_{ij} = x_{kj}$ (for *i*-th and *k*-th object).

References

Gatnar, E., Walesiak, M. (Eds.) (2004), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych [Multivariate statistical analysis methods in marketing research]*, Wydawnictwo AE, Wroclaw, p. 43.

Kaufman, L., Rousseeuw, P.J. (1990), *Finding groups in data: an introduction to cluster analysis*, Wiley, New York, p. 28.