Package 'bfcluster'

November 24, 2025

Title Buttler-Fickel Distance and R2 for Mixed-Scale Cluster Analysis

Version 1.0.0

Description Implements the distance measure for mixed-scale variables proposed by Buttler and Fickel (1995), based on normalized mean pairwise distances (Gini mean difference), and an R2 statistic to assess clustering quality.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.3

Depends R (>= 4.0.0)

NeedsCompilation no

Author Moritz Schäfer [aut, cre]

Maintainer Moritz Schäfer <moritz1.schaefer@uni-a.de>

Repository CRAN

Date/Publication 2025-11-24 09:30:22 UTC

Contents

`bf_R2` *R^2 for Cluster Solutions after Buttler & Fickel (1995)*

Description

Computes the proportion of explained distance variation (R^2) for a given clustering solution using a distance matrix derived from the Buttler-Fickel distance. The statistic reflects how well the clustering partitions the total pairwise distance structure.

Usage

```
bf_R2(D, cluster)
```

Arguments

D	A distance object of class dist, usually computed via buttler_fickel_dist().
cluster	An integer or factor vector of cluster assignments, typically obtained from cutree() or another clustering method.

Details

The R² is defined as:

$$R^2 = 1 - \frac{D_{\text{within}}}{D_{\text{total}}}$$

where D_{total} is the sum of all pairwise distances and D_{within} is the sum of distances within clusters.

Value

A numeric value between 0 and 1 indicating the proportion of explained distance variation. Higher values represent better cluster fit.

Examples

```
df <- data.frame(
  sex    = factor(c("m","f","m","f")),
  height = c(180, 165, 170, 159),
  age    = c(25, 32, 29, 28)
)

types <- c("nominal", "metric", "metric")

D <- buttler_fickel_dist(df, types)
hc <- hclust(D)
cl <- cutree(hc, k = 2)

bf_R2(D, cl)
```

---

buttler_fickel_dist *Buttler-Fickel Distance Matrix*

Description

Computes a distance matrix following Buttler & Fickel (1995) for mixed-scale variables. Each variable-specific distance matrix is normalized by its mean pairwise distance (Gini mean difference), ensuring equal contribution of all variables to the overall distance.

Usage

```
buttler_fickel_dist(df, types)
```

Arguments

df	A data.frame where rows are cases and columns are variables.
types	A character vector of the same length as ncol(df), indicating the scale level of each variable. Allowed values are "metric", "ordinal", or "nominal".

Value

An object of class dist.

Index