Calculate the inconsistency of a hclust object.

inconsistent

Description

Calculates the inconsistency of a 'hclust' object and splits the data into clusters using a inconsistency-based cut-off.

Usage

inconsistent(hclust_obj, cut_point = NULL)

Arguments

hclust_obj the input data - see details.

cut_point a value of inconsistency at which to split the clusters into groups.

Details

This function calculates the inconsistency of a hclust object. This provides the same functionality as the inconsistent function in Matlab and the 'scipy" library in Python. The inconsistency can be calculated as

$$\frac{h_i - \bar{h}}{s}$$

where h_i is the height of the given subtree; and \bar{h} and s are respectively the mean and standard deviation of the height of the subtree and any non-leaf subtrees. Note that the mean and standard deviationare calculated over 1, 2, or 3 heights depending on whether each subtree has 0, 1, or 2 non-leaf subtrees. By definition, a subtree where both of its subtrees are leaves has an inconsistency of 0.

Value

The original object of class hclust, with additional components containing the calculated inconsistency at each merge point and, if a cut-point is specified, the resulting clusters.

inconsistency The calculated inconsistency at each merge point.

clusters A data frame with 2 variables: the label and the assigned cluster. (Only

present if a cut_point is specified.)

See Also

hclust

Examples

```
hc <- hclust(dist(USArrests), "ave")
hc_obj <- inconsistent(hc, 1)

# the first 6 merges all involve leaves, so the inconsistency is 0.
head(with(hc_obj, cbind(merge, height, inconsistency)))

# the last 6 merges are all non-leaves, so the inconsistency can be calculated tail(with(hc_obj, cbind(merge, height, inconsistency)))

# Each element of the tree is assigned to a cluster.
head(hc_obj$clusters)
table(hc_obj$clusters$cluster)</pre>
```