factorMerger
Cheat Sheet

Agnieszka Sitko [aut, cre]
Przemysław Biecek [aut, ths]
University of Warsaw

MI

Basics

The aim of **factorMerger** is to provide set of tools to support results from post hoc comparisons. *Post hoc* testing is an analysis performed after running *ANOVA* to examine differences between group means (of some response numeric variable) for each pair of groups (groups are defined by a factor variable).

This project arose from the need to create a method of post hoc testing which gives the hierarchical interpretation of relations between groups means. Thereby, for a given significance level we may divide groups into nonoverlapping clusters.

1.TODO: here describe more details

Example

```
library(factorMerger)
fmAll <- mergeFactors
    (response_variable,
    factor_variable,
    method = "hclust",
    family = "binomial")
print(fmAll)
plotTree(fmAll,
    panel = "all",
    responsePanel = "tukey")
```
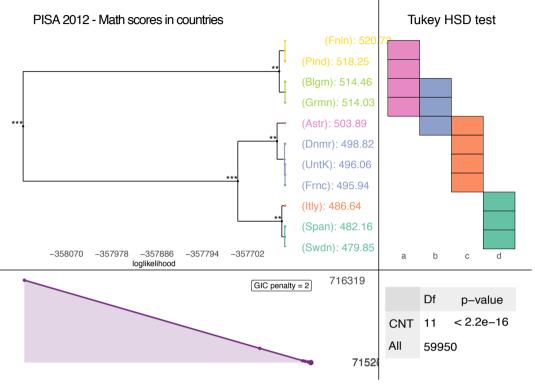
Four panels to see it all

A. The Merging Paths Plot panel
Shows the hierarchical structure of similarity between groups.
Stars presents how significant are differences between two clusters.
Height of the join corresponds to the likelihood of the combined model.

B. The response panel
Shows graphical summaries for group separately.
Use the responsePanel argument to select the desired summary.

PISA 2012 - Math scores in countries

Tukey HSD test

	Df	p-value
CNT	11	< 2.2e−16
All	59950	

C. The GIC panel
Shows the *Generalized Information Criterion* for all models in MPP panel.
Default *penalty=2* corresponds to the AIC criterion.
Colours in the MPP panel corresponds to the optimal model.

D. The summary panel
Shows the no. groups, no. observations, and the p-value for global hypothesis that parameter of interest is equall in all groups.
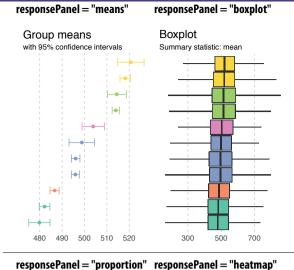
Level fusing plot

The top-left plot shows level fusing paths (merging paths). With arguments **family=**, **show=**, **fuse=**, **spacing=**, one can select how to merge factors and what shall be presented on OX/OY axes.
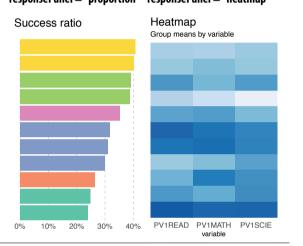
Argument	Summary
panel = "all"	All panels
panel = "response"	Panels A and B
panel = "GIC"	Panels A and C
panel = "tree"	Only panel A

Argument	Summary
fuse = "all2all"	Compare all pairs of groups
fuse = "nearby"	Compare nearby groups
fuse = "cluster"	DMR4glm algorithm

Argument	Summary
family = "gaussian"	For 1/n-dimensional Gaussian
family = "binomial"	For binomial regression
family = "survival"	For Cox regression

The response panel summaries

responsePanel = "means"

Group means
with 95% confidence intervals

responsePanel = "boxplot"

Boxplot
Summary statistic: mean

responsePanel = "proportion"

Success ratio

responsePanel = "heatmap"

Heatmap
Group means by variable

responsePanel = "tukey"

Tukey HSD test

responsePanel = "frequency"

Groups frequencies

responsePanel = "profile"

Profile plot
Variable means ranks

responsePanel = "survival"

Survival plot
Survival curves for coxph model