

# factorMerger

## Cheat Sheet

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



### Introduction

How to check if averages are different among k groups? Use ANOVA!

How to visualise how these groups are different? Use **factorMerger**!

The aim of **factorMerger** is to provide informative and easy to understand visualisations of *post-hoc* comparisons. It gives consistent and non-overlapping adaptive fusing of groups based on likelihood ratio test (LRT). The package **factorMerger** works for wide spectrum of families like Gaussian, binomial or survival.

Results from the adaptive fusing are presented with the *Merging Paths Plots* - a hierarchical representation of LRT-based distances among groups.

In addition, the *Generalized Information Criterion* (GIC) is presented for fused models. This criterion may be used to choose the optimal segmentation of groups.

Graphical summary of the variable of interest in each group is presented in the right panel.

Find more in <https://arxiv.org/abs/1709.04412>

### Example

#### Library (factorMerger)

```
fmAll <- mergeFactors(
  response = pisaEuro$math,
  factor = pisaEuro$country,
  method = "fast-adaptive",
  family = "gaussian")
```

```
print(fmAll)
```

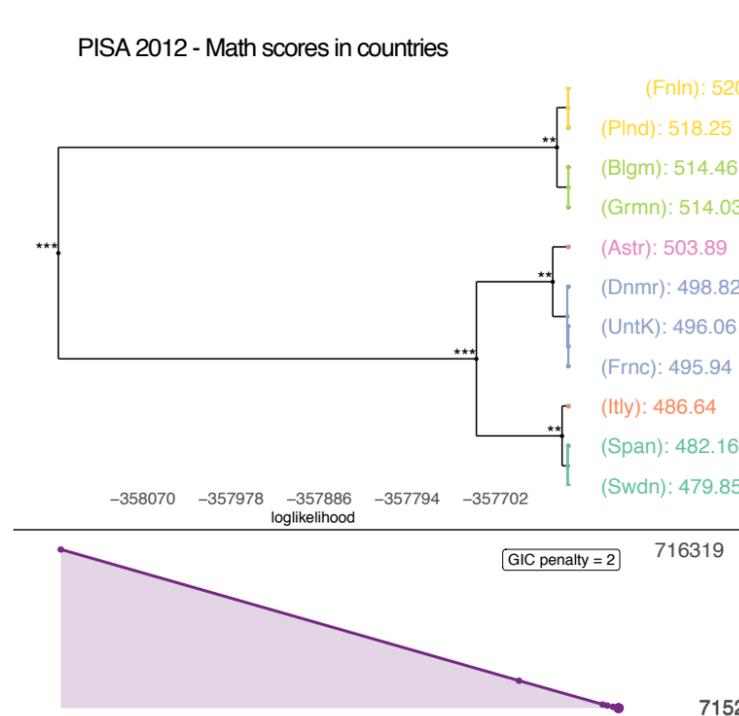
```
plot(fmAll,
  panel = "all",
  responsePanel = "tukey")
```

## Set of tools to support post-hoc segmentation of groups

### Four panels to see it all

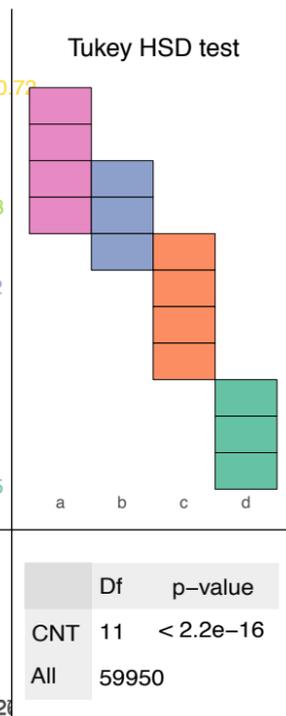
#### A. The Merging Paths Plot panel

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.



#### 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 equal in all groups.

### Key arguments of factorMerger

The **factorMerger** package is highly customisable.

The **panel** argument selects panels to be plotted.

The **nodesSpacing** argument selects how MPP nodes shall be spaced.

The **responsePanel** argument selects which summary shall be presented in panel B.

The **responsePanelPalette** argument selects the colouring palette.

The **colorClusters** argument selects if segments of groups shall be colour coded.

The **penalty** argument selects penalty for GIC criterion.

The **method** argument determines how pairs of groups are considered for grouping. The most computational demanding is *adaptive*. For larger number of groups one may consider only pairs of groups with closest averages or use the *fast-fixed* heuristic (fastest solution).

Argument	Summary
method = "adaptive"	Compare all pairs of groups
method = "fast-adaptive"	Compare nearby groups
method = "fixed"	Delete or merge heuristic
method = "fast-fixed"	Reduced DMR method

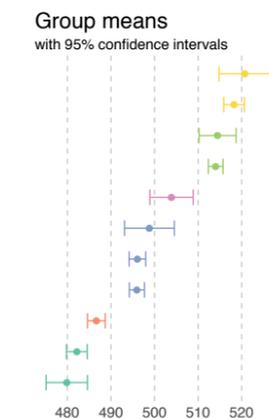
The **family** argument determines what models are to be compared with likelihood ratio test. Either linear models, generalised linear models or Cox regression models.

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

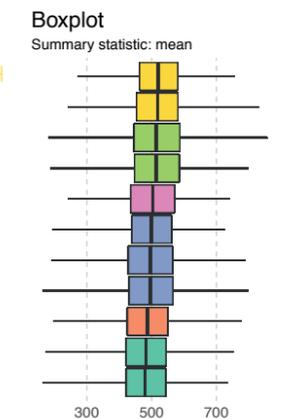
Argument	Summary
panel = "all"	All panels
panel = "response"	Panels A and B
panel = "GIC"	Panels A and C
panel = "tree"	Only panel A

### The response panel summaries

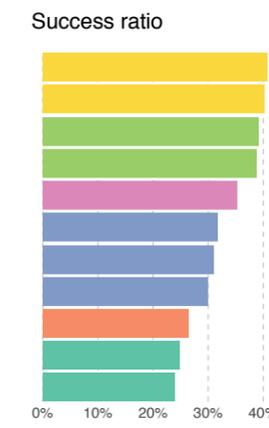
#### responsePanel = "means"



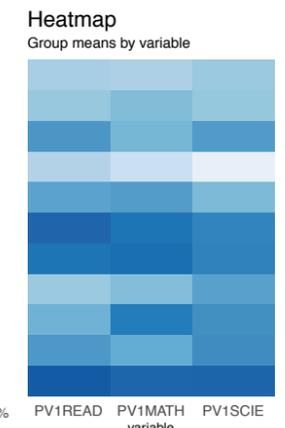
#### responsePanel = "boxplot"



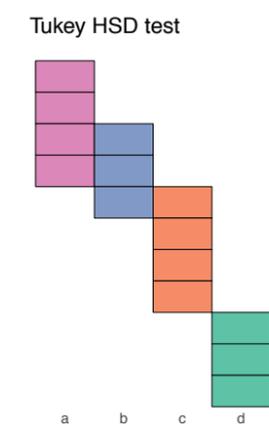
#### responsePanel = "proportion"



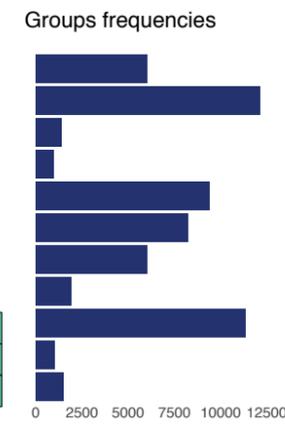
#### responsePanel = "heatmap"



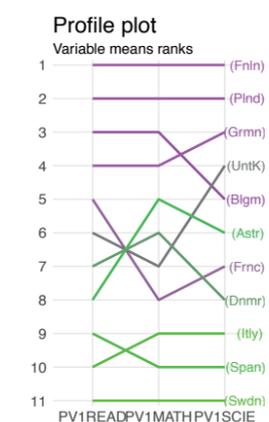
#### responsePanel = "tukey"



#### responsePanel = "frequency"



#### responsePanel = "profile"



#### responsePanel = "survival"

