Functions O-tree

Pedro Salas Rojo (with Domenico Moramarco and Paolo Brunori)

July 2024

Documentation for the R functions used to estimate the Opportunity tree (O-tree) and the reward and compensation scores Moramarco et al. (2024). Please, note that these functions merely provide guidance for initial practitioners, and it is the researcher's responsibility to appropriately produce the estimates.

1 Libraries

Libraries are updated on the 1st of July, 2024.

Library	Resource
partykit	https://cran.r-project.org/web/packages/partykit/index.html
tidyverse	https://cran.r-project.org/web/packages/tidyverse/index.html
pwr	https://cran.r-project.org/web/packages/pwr/index.html
data.table	https://cran.r-project.org/web/packages/data.table/index.html
tm	https://cran.r-project.org/web/packages/tm/index.html
dineq	https://cran.r-project.org/web/packages/dineq/index.html
strucchange	https://cran.r-project.org/web/packages/strucchange/index.html

2 Code

2.1 O-tree

o_tree(data, model, types, outcome, circum, weights, alp = 0.05, mu = 0, minbucket = 100, merge = c("ctree", "t.test"), try.new = TRUE)

- data: name of dataframe.

- model: formula, e.g., dependent   independents.

- types: name of the original type partition, e.g. "types".

- outcome: name of the outcome variable, e.g. "income".

- circum: vector of circumstances to check, e.g. c("ethnicity","sex")

- weights: vector of weights, e.g. "weights". If no weights in your data, substitute by a vector of value 1.

- alp: confidence level used in tests. Default = 0.05

- mu: difference in means to be tested. Default = 0.

- minbucket: minbucket parameter in "partykit", see Brunori et al. (2023). Default = 100.

- merge: algorithm to merge types. It can be a ctree or a t-test.

- : try.new: set to TRUE (default) if, after merging, you want the o-tree to find new types.

The function returns the original dataset (including the original type partition) with two new columns:

- "m.types" showing the types after merging, first step of the O-tree.

- "o.types" showing the types after merging and trying a further split, second step of the O-tree.

2.2   Reward score

reward_score(data, types, outcome, alp = 0.01, mu = 0, eps = 200)

- data: name of dataframe.

- types: name of the original type partition, e.g. "types".

- outcome: name of the outcome variable, e.g. "income".

- alp: confidence level used in tests. Default = 0.01

- mu: difference in means to be tested. Default = 0.

- eps: epsilon to compute power. Default = 200.

The reward score function deliver a "results" dataframe with the main results (types, pvalues, power,...) to compute the score. It also delivers the score.

2.3   Compensation score

compen_score(data, types, outcome, model, circum, alp = 0.01, mu = 0, eps = 200)

- data: name of dataframe.

- types: name of the original type partition, e.g. "types".

- outcome: name of the outcome variable, e.g. "income".

- model: formula, e.g., dependent   independents.

- circum: vector of circumstances to check, e.g. c("ethnicity","sex")

- alp: confidence level used in tests. Default = 0.01

- mu: difference in means to be tested. Default = 0.

- eps: epsilon to compute power. Default = 200.

The compensation score function deliver a "results" dataframe with the main results (types, pvalues, power,...) to compute the score. It also delivers the score.

References

Brunori, P., Hufe, P., and Mahler, D. (2023). The roots of inequality: Estimating inequality of opportunity from regression trees and forests. *The Scandinavian Journal of Economics*, 125(4):900–932.

Moramarco, D., Brunori, P., and Salas-Rojo, P. (2024). Biases in inequality of opportunity estimates: measures and solutions. *mimeo*, 1(1).