

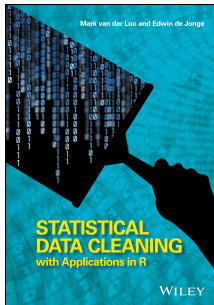
Adjusting numerical values

Mark van der Loo and Edwin de Jonge

Statistics Netherlands Research & Development

Twitter: @markvdloo @edwindjonge

Tokyo | Institute of Statistical Mathematics | 2020



Try the code

```
03valid/adjusting.R
```



Adjusting numerical values

Minimally adjust values so that they conform to rules after imputation.



Imputation

- Most imputation methods do not take the data restrictions/rules into account.
- This means that valid data can become invalid after missing values have been imputed.



Successive projection algorithm

Idea

Alter (imputed) values in a record \mathbf{x} *as little as possible* to satisfy all restrictions.

As little as possible?

The minimal Euclidian distance between the original \mathbf{x} and the adjusted record \mathbf{x}^* .

$$\mathbf{x}^* = \min_{\mathbf{x}} (\mathbf{x}^* - \mathbf{x})'(\mathbf{x}^* - \mathbf{x})$$

Successive Projection Algorithm (sketch)

Project \mathbf{x} on each (in)equality restriction sequentially and iteratively until convergence.

Hildredth (1957) *Naval Research Logistics* 4 79–85



Extension: weighted distance

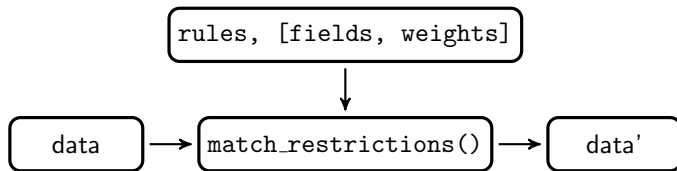
$$\mathbf{x}^* = \min_{\mathbf{x}} (\mathbf{x}^* - \mathbf{x})' \mathbf{W} (\mathbf{x}^* - \mathbf{x})$$

Property

If $W_{ij} = \delta_{ij} x_j^{-1}$, then the ratios between altered variables are preserved to $\mathcal{O}(1)$.

Pannekoek & Zhang (2015) *Survey Methodology* **41** 127–144; SDCR §10.11





Assignments

- load “03valid/errors_located.csv” into `errors_located`
- load “03valid/imputed.csv” into `imputed`.
- use `confront` to find out how many values are invalid in `imputed` and make a plot of the object
- Use `is.na` to store all NA values of `errors_located` into `adjust`
- apply `rspa::match_restrictions` to the data and use the `adjust` argument: we are restricting adjustments to the data that are imputed.
- use `confront` to find out how many values are invalid and make a plot of the object

