



Mean / Median imputation per group

Mean / Median imputation: definition

- Replacement value = variable mean or median
- Suitable numerical variables.

Mean / Median imputation per group

- Replacement value = variable mean or median based on some grouping.
- Suitable numerical variables.
- Grouping variable should be categorical or discrete.

Mean imputation per group

Gender	Height
1	100
1	90
1	
1	70
1	80
1	
0	60
0	60
0	70
0	
0	70

Mean = 85



Mean = 65

Gender	Height
1	100
1	90
1	
1	70
1	80
1	
0	60
0	60
0	70
0	
0	70

Mean imputation per group

Gender	Height
1	100
1	90
1	
1	70
1	80
1	
0	60
0	60
0	70
0	
0	70

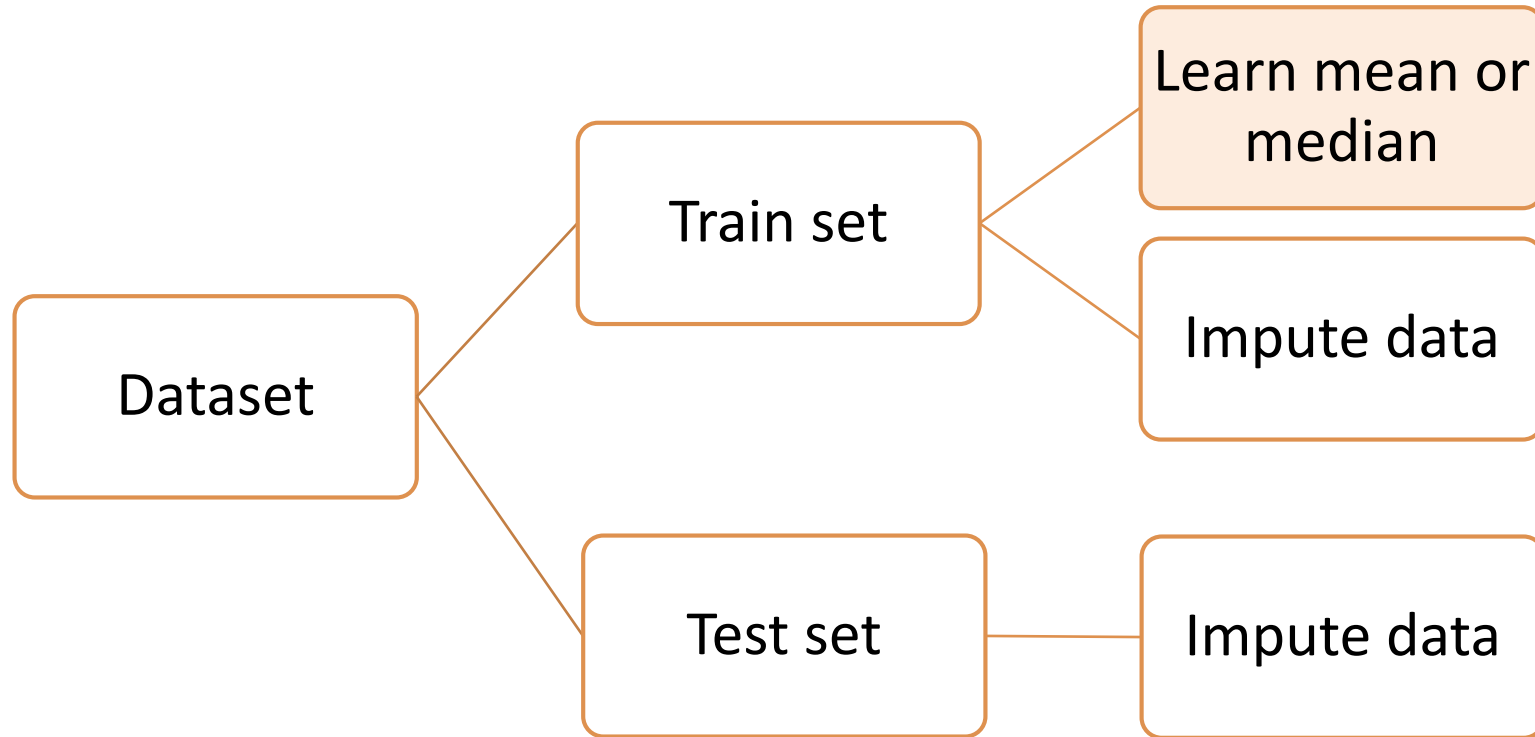
Mean = 85



Mean = 65

Gender	Height
1	100
1	90
1	85
1	70
1	80
1	85
0	60
0	60
0	70
0	65
0	70

Correct workflow



The mean or the median are “learned parameters”, like the coefficients of a linear model, or the splits of a tree.

■ ■ Why would we use
● this method?



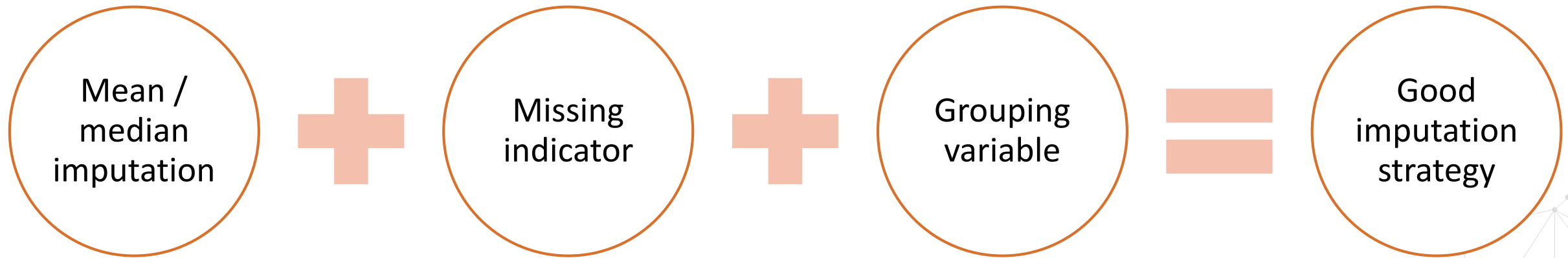


Why?

- We'd like a better representation of the missing data.
- The variable distribution differs based on the grouping variable.



Good imputation strategy



THANK YOU

www.trainindata.com