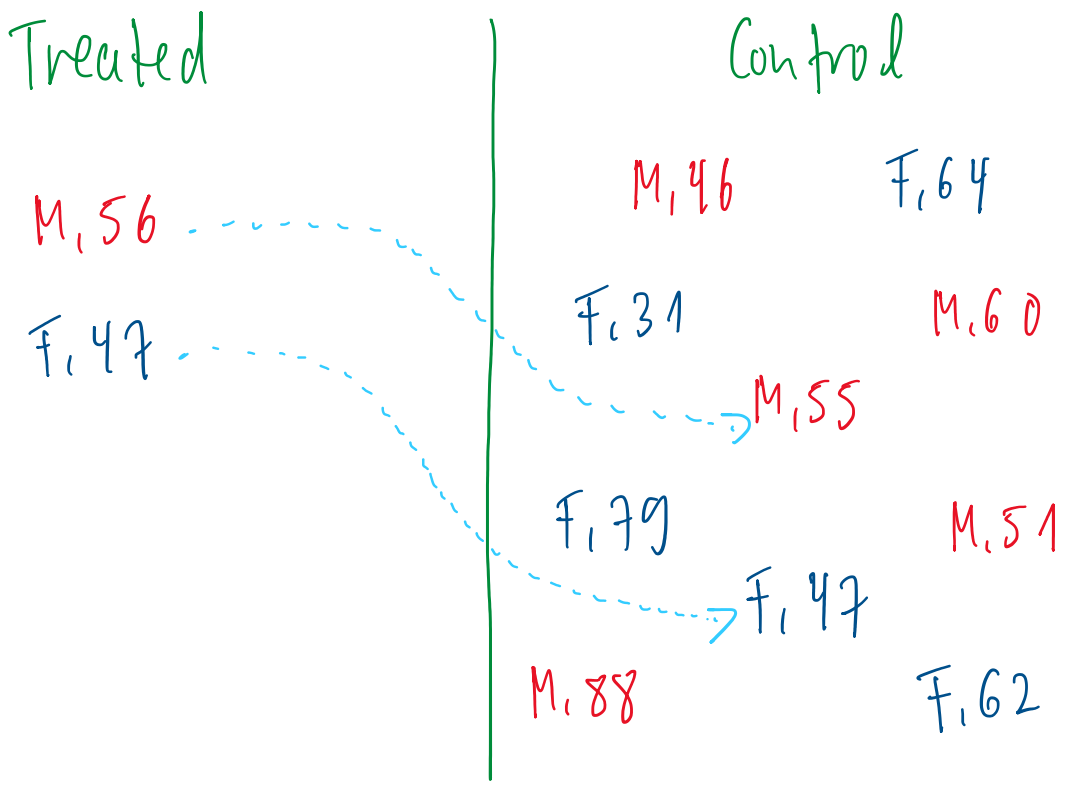
Matching (overview)

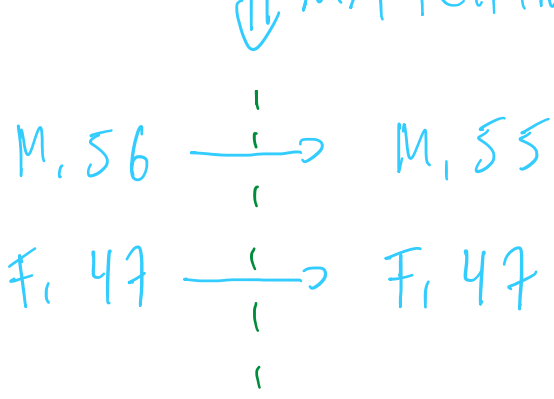
03.05.21 10:02

- imbalance on the covariate X btw. different treatment groups
- match treated subject with control subject based on similar covariate values X
- matching gets difficult if X is high-dimensional
- goal: to achieve stochastic balance btw. the dist. of covariates in each treatment group

Example: - assume we have $X \in \mathbb{R}^2$ with X_1 .. age X_2 ... sex (M|F)



⇓ MATCHING GOAL: TO CONTROL FOR X



} allows to estimate the causal effect of the treatment on the treated (bc we adjust sample according to treated population)

⇓ We are making the dist. of X in the control group look like that in the treated population

Fine balance concept:

- Sometimes it is difficult to find great matches
  ↳ we might be willing to accept non-ideal matches if treated and control group end up with the same marginal dist. of covariates → FINE BALANCE

Example:

match 1: A=1, M, 40 – A=0, F, 45
match 2: A=1, F, 45 – A=1, M, 40

} "bad matches" but avg. (marginal dist. in each group for each covariate are same

example of good fine balance even though matches are "bad".

group A=1:
(1) avg. age 42.5
(1) 50% female

group 2:
(1) avg. age 42.5
(1) 50% female

# of matches:

(1) pair matching (1-to-1): match exactly one control to every treated subject
(2) many-to-one: match fixed number (K) controls to every treated subject (e.g. 5-to-1)
(3) variable: sometimes match 1, other times match more than 1, control to treated subjects
  ↳ if there are multiple "good" matches, we may want to use them!
  ↳ 1-to-N matching, with N ∈ ℕ being flexible