News from Cyperspace

Christian Torp-Pedersen

September 24, 2022

LPR

```
# Good old LPR2 until 2018
proc1 <- importSAS('X:/data/rawdata_hurtig/702773/opr.
    sas7bdat',
          where=......
          keep=c("pnr","opr","odto"),date.vars = "odto
    ")
# LPR2 remnants 2019
proc2_2019 <- importSAS('X:/data/rawdata_hurtig/703775/
    2022/t_sksube.sas7bdag',
          where=.....,
          keep=c("pnr","c_opr","d_odto","recnum"))
setnames(proc2_2019,c("c_opr","d_odto"),c("opr","odto")
    )
```

LPR

```
#Great new LPR3
proc3 <- importSAS('X:/data/workdata_hurtig
    /703775/2022/procedurer.sas7bdat',
            where=....)
#make a filter, LPR3 is HUGE
filt <- unique(proc3[,.(kontakt_id)])
#Filtered version of adm3
adm3 <- importSAS('X:/data/rawdata_hurtig/703775/2022/
    kontakt.sas7bdat',,
          keep=c("kontakt_id","pnr"),filter=filt)
proc3 <- merge(proc3,adm3,by="kontakt_id")
proc3[,odto:=as.Date(starttidspunkt)]
proc3 <- proc3[,.(pnr,odto)]
proc <- rbind(proc1,proc2_2019,proc3,fill=TRUE)
```

TARGETS

`R/functions.R` contains our custom user-defined functions. (See the [functions chapter](#) for a discussion of function-oriented workflows.)

```r
# R/functions.R
get_data <- function(file) {
  read_csv(file, col_types = cols()) %>%
    filter(!is.na(Ozone))
}

fit_model <- function(data) {
  lm(Ozone ~ Temp, data) %>%
    coefficients()
}

plot_model <- function(model, data) {
  ggplot(data) +
    geom_point(aes(x = Temp, y = Ozone)) +
    geom_abline(intercept = model[1], slope = model[2])
}
```
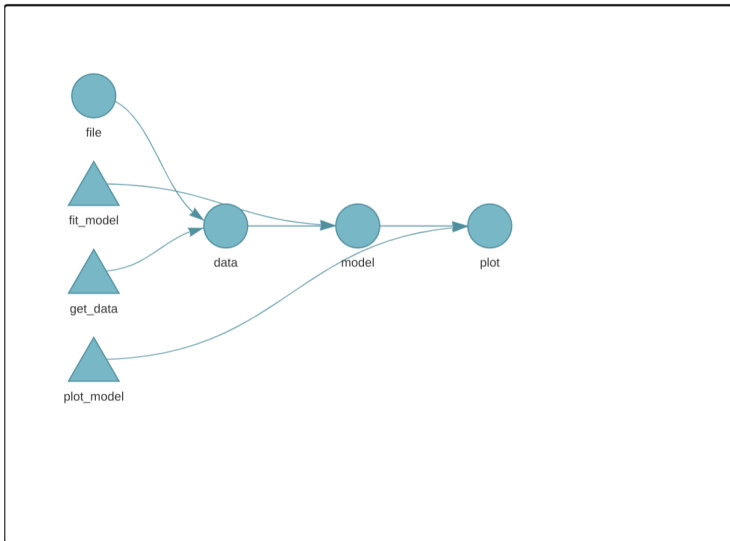
```r
# _targets.R file
library(targets)
source("R/functions.R")
tar_option_set(packages = c("readr", "dplyr", "ggplot2"))
list(
  tar_target(file, "data.csv", format = "file"),
  tar_target(data, get_data(file)),
  tar_target(model, fit_model(data)),
  tar_target(plot, plot_model(model, data))
)
```

file

fit_model

get_data

data

model

plot

plot_model

```
tar_make()
#> • start target file
#> • built target file
#> • start target data
#> • built target data
#> • start target model
#> • built target model
#> • start target plot
#> • built target plot
#> • end pipeline: 1.007 seconds
```

```
tar_make()
#> ✔ skip target file
#> ✔ skip target data
#> ✔ skip target model
#> ✔ skip target plot
#> ✔ skip pipeline: 0.095 seconds
```

2.6.1 Change code

If you change one of your functions, the targets that depend on it will no longer be up to date, and `tar_make()` will rebuild them. For example, let's increase the font size of the plot.

```
# Edit functions.R...
plot_model <- function(model, data) {
  ggplot(data) +
    geom_point(aes(x = Temp, y = Ozone)) +
    geom_abline(intercept = model[1], slope = model[2]) +
    theme_gray(24) # Increased the font size.
}
```
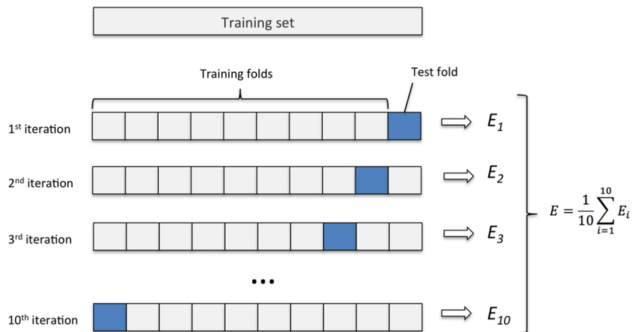
```
tar_visnetwork()
```
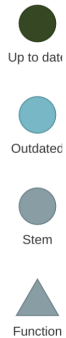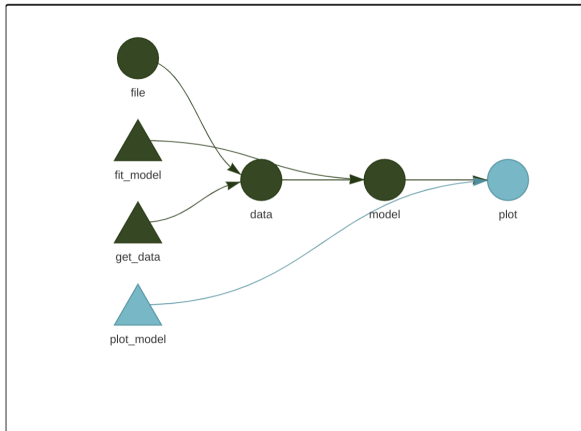
Counterfactual

Causal Inference

- Parametric G-formula
- Propensity matching
- Inverse probability weighting

Superlearner

Propensity adjustment

Treated

$\Pr[A = 1 | X = \text{adult, male}] = {}^{1}/_{5}$
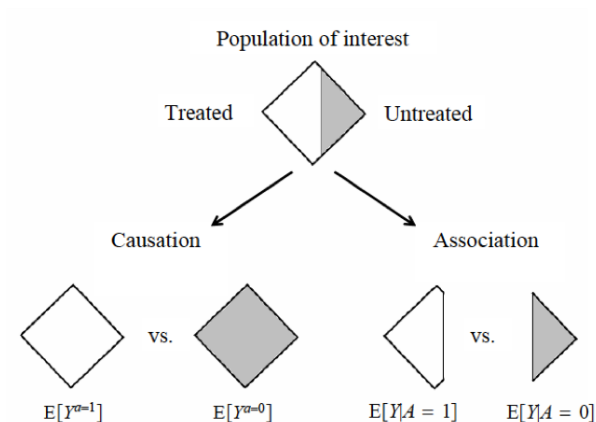
$w_i = 5$

Untreated

$\Pr[A = 0 | X = \text{adult, male}] = {}^{4}/_{5}$

$w_i = {}^{5}/_{4}$

Our pseudo-population includes a similar number of adult males (note the duplicated Homer and the hybrid untreated), so when comparing the groups — the effect of adult-males will cancel it-self and we'll be left only with the effect of the treatment.

G-estimation

TMLE - LTMLE

Targeted Maximum Likelihood Estimation

- ▶ Step-1 Outcome versus covariates using superlearner
- ▶ Step-2 Targeting using propensity of treatment
- ▶ Double robust G-estimation of target parameter

Target Parameter flexible such as calculating the benefit/risk of 5 ears of treatment with X versus no treatment

PhD course: Targeted Register Analysis

Dates: 12, 13, 14 and 15 December 2022, all days 8.00-15.00
Course location: CSS
Registration: Please register before 7 November 2022

- Analysing Danish register data
- The roadmap of targeted statistical learning
- The transition from traditional epidemiological tools (cohort followup studies, case-control studies) which produce hazard ratios or odds ratios to average treatment effects defined in a dynamic causal framework
- Machine learning (random forests/recursive neural networks)
- Longitudinal minimum loss estimation (LTMLE)
- Use the R-package targets to setup and organize a reproducible analysis

Teachers

Christian Torp-Pedersen, Nothern Sealands Hospital and Section of Biostatistics, Department of Public Health, University of Copenhagen

Marvin Wright, Leibniz Institute for Prevention Research and Epidemiology, Bremen, Germany

Zeyi Wang, Postdoctoral scholar, Division of Biostatistics, University of California, Berkeley

Andrew Mertens, postdoc, Division of Biostatistics ,University of California, Berkeley

Thomas A. Gerds, Section of Biostatistics, Department of Public Health, University of Copenhagen