

# Second Modeling Run

*Martin Skarzynski*

*2018-04-20*

```
library(here)
library(readr)
library(dplyr)
library(tidyr)
library(survey)
library(purrr)
library(furrr)
plan(multiprocess)
source("kod/get_modelstats.R")

#Variables to be included in the model
chsn <- c(
  "PERMTH_INT", #time
  "canc_mort", #event
  "SDPPSU6", #PSU
  "SDPSTRA6", #Stratification
  "WTPFQX6", #Weights
  "HSAITMOR", #Age in months at interview (screener)
  "DMAETHNR", #Ethnicity
  "HSSEX", #Biological sex
  "HAT16", #In the past month, did you lift weights
  "HAK9", # times per night you get up to urinate
  "HAQ1", #Describe natural teeth: excellent...poor
  "HAR1", #Have you smoked 100+ cigarettes in life
  "HAT2", #In the past month, did you jog or run
  "HAT18", #In the past month, any other exercises, sports
  "HAB1" #Would you say your health in general is excellent, very good, good, fair, or poor?
)

#Variables to be removed from the model
remv <- c(
  "HAN9", #remove age variables
  "HAQ7",
  "HAT29",
  "HAJ0",
  "WTPXRP2", #remove unneed weight variables
  "WTPQRP21",
  "WTPQRP27",
  "WTPQRP43"
)

path <- 'dat/3-clean-complete-cases.rds'

# the modelstats function returns a dataframe
# after taking in the following arguments
#seed,
#n_random_vars,
#datafile_path,
#chosen_vars,
```

```

#remove_vars
map_sizes <- function(seed){
  map2_dfr(.x = seed,
    .y = seq(10), # number of random variables
    ~get_modelstats(seed=.x,
      n_random_vars=.y,
      datafile_path=path,
      chosen_vars=chsn,
      remove_vars=remv))
}
#save an object with 1000 models
map_dfr(.x = seq(50), # number of seeds
  .f = ~map_sizes(seed=.x)) %>%
write_rds(here("dat/5-model-second-run.rds"))

```