sampling

Jules Brable

2022-12-12

SETTING ENVIRONEMENT

```
rm(list = ls())

library(data.table)
library(tidyverse)

source(here::here("R/statdesc_functions.R"))
source(here::here("R/sampling_functions.R"))

list.files(path = "data")
```

```
## character(0)
```

```
## import data ##
# loading all 2021 data in order to compare
tab2021 <- load_data("nat2021us.csv")

# designing our sample #

# issue with the column names #
ndiff1 <- setdiff(
  (sample_df("nat2019us.csv", n = 1, segment_treat = F) %>% colnames()),
  (sample_df("nat2021us.csv", n = 1, segment_treat = F) %>% colnames())
)

ndiff2 <- setdiff(
  (sample_df("nat2021us.csv", n = 1, segment_treat = F) %>% colnames()),
  (sample_df("nat2019us.csv", n = 1, segment_treat = F) %>% colnames())
)

to_remove <- c(ndiff1, ndiff2)

n <- 300000
s2021 <- sample_df("nat2021us.csv",  n, remove = to_remove)
s2020 <- sample_df("nat2020us.csv", n, remove = to_remove)
s2019 <- sample_df("nat2019us.csv", n, remove = to_remove)
s2018 <- sample_df("nat2018us.csv", n, remove = to_remove)
```

var	N	U	Y	X
Infertility Treatment Used	0.977	0.001	0.022	NA
Fertility Enhancing Drugs	0.013	0.002	0.008	0.977
Asst. Reproductive Technology	0.006	0.002	0.015	0.977

var	N	Y	U	X
Infertility Treatment Used	0.500	0.500	NA	NA
Fertility Enhancing Drugs	0.013	0.008	0.002	0.977
Asst. Reproductive Technology	0.006	0.015	0.002	0.977

```
mixsample <- s2021 %>%
  bind_rows(s2020) %>%
  bind_rows(s2019) %>%
  bind_rows(s2018)

pma <- c("Infertility Treatment Used",
         "Fertility Enhancing Drugs",
         "Asst. Reproductive Technology")
```

STAT DESC

- Sur les 3669928 données de 2021, on a:

```
tab2021 %>%
  get_prop(rf_inftr) %>%
  bind_rows(tab2021 %>%
              get_prop(rf_fedrg)) %>%
  bind_rows(tab2021 %>%
              get_prop(rf_artec)) %>%
  mutate(var = pma) %>%
  relocate(var) %>%
  kableExtra::kbl() %>%
  template()
```

- Sur les 1200000 données de notre échantillon (2019 + 2020 + 2021), on a: