Sampling and calibration weights

We describe the construction of the sampling and calibration weights of the COVID-19 social monitor online panel.

Author: André Moser, Epidemiology, Biostatistics and Prevention Institute, University of Zurich

Date: May 02, 2020

```r
library(tidyverse)
library(survey)
library(INLA)

### Read data
data <- read_csv(paste0(main.path, "w1_selected.csv"))
age_gender_lregion <- read_csv(paste0(main.path, "age_gender_lregion.csv"))
age_educ_gregion <- read_csv(paste0(main.path, "age_educ_gregion.csv"))

### Bayesian multilevel model with non-informative priors (INLA)
inla.mod <- inla(n ~ f(agecat, model = "iid")+f(gender, model = "iid")+f(lregion, model = "iid"),
                 data = age_gender_lregion, family="binomial", control.predictor = list(link = 1),
                 Ntrials=age_gender_lregion$num)

### Join predictions to data
age_gender_lregion$pred <- inla.mod$summary.fitted.values[,1]
data <- left_join(x=data, y=age_gender_lregion,
                  by=c("agecat"="agecat", "lregion"="lregion", "gender"="gender"))
### Create sampling weights
data$sampling_weight <- round(1/data$pred,0)

### Aggregate education because of small cell freq
data$educ2 <- data$educ
data$educ2[data$educ2==1] <- 2

age_educ_gregion$educ2 <- age_educ_gregion$educ
age_educ_gregion$educ2[age_educ_gregion$educ2==1] <- 2

w1_agg <- data %>% group_by(agecat, gregion, educ2) %>% summarise(n=n())

bfs_agg <- age_educ_gregion %>% group_by(agecat, reg, educ2) %>% summarise(num=sum(n))

data_agg <- left_join(x=bfs_agg, y=w1_agg,
                      by=c("agecat"="agecat", "reg"="gregion", "educ2"="educ2"))

## Use survey library
wdesign <- svydesign(id=~1, strata=~agecat+gender+lregion, weights=~sampling_weight, data=data)
data_svy <- data_agg %>% select(gregion=reg, agecat, educ2, Freq=num)
wdesign <- postStratify(wdesign, strata=~gregion+educ2+agecat, population=data_svy )
wdesign <- trimWeights(wdesign, upper=10000)
```

```r
## Calibrated weights
data$ps_weight <- round(1/summary(wdesign)$prob)

summary(data$sampling_weight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     917    3429    3452    3496    3817    5940
```

```r
summary(data$ps_weight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     495    2180    2741    3063    3464   10000
```