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: August 27, 2020

```
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.
##     918    3427    3451    3493    3815    5902
```

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     490    2178    2735    3060    3459   10000
```

```r
## Construction of non-response weights

mod <- glm(w2 ~ factor(agecat)+factor(gender)+factor(lregion)+factor(educ)
           +factor(partner)+factor(work), data=data, family=binomial())

data$nr_weight_w2 <- round(1/predict(mod, type="response"),3)
data$nr_weight_w2[data$w2==0] <- round(1/(1-predict(mod, type="response")[data$w2==0]),3)

summary(data$nr_weight_w2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.171   1.293   1.332   1.999   1.505   6.035
```

```r
mod <- glm(w3 ~ factor(agecat)+factor(gender)+factor(lregion)+factor(educ)
           +factor(partner)+factor(work), data=data, family=binomial())

data$nr_weight_w3 <- round(1/predict(mod, type="response"),3)
data$nr_weight_w3[data$w3==0] <- round(1/(1-predict(mod, type="response")[data$w3==0]),3)

summary(data$nr_weight_w3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.087   1.295   1.329   2.001   1.475  11.128
```

```r
mod <- glm(w4 ~ factor(agecat)+factor(gender)+factor(lregion)+factor(educ)
           +factor(partner)+factor(work), data=data, family=binomial())

data$nr_weight_w4 <- round(1/predict(mod, type="response"),3)
data$nr_weight_w4[data$w4==0] <- round(1/(1-predict(mod, type="response")[data$w4==0]),3)

summary(data$nr_weight_w4)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.038   1.117   1.176   2.008   1.259  20.978
```

```r
mod <- glm(w5 ~ factor(agecat)+factor(gender)+factor(lregion)+factor(educ)
           +factor(partner)+factor(work), data=data, family=binomial())

data$nr_weight_w5 <- round(1/predict(mod, type="response"),3)
data$nr_weight_w5[data$w5==0] <- round(1/(1-predict(mod, type="response")[data$w5==0]),3)

summary(data$nr_weight_w5)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##   1.081   1.156   1.233   2.002   1.304  12.234
```

```
mod <- glm(w6 ~ factor(agecat)+factor(gender)+factor(lregion)+factor(educ)
           +factor(partner)+factor(work), data=data, family=binomial())

data$nr_weight_w6 <- round(1/predict(mod, type="response"),3)
data$nr_weight_w6[data$w6==0] <- round(1/(1-predict(mod, type="response")[data$w6==0]),3)

summary(data$nr_weight_w6)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.099   1.197   1.281   2.002   1.395  10.363
```

```
mod <- glm(w7 ~ factor(agecat)+factor(gender)+factor(lregion)+factor(educ)
           +factor(partner)+factor(work), data=data, family=binomial())

data$nr_weight_w7 <- round(1/predict(mod, type="response"),3)
data$nr_weight_w7[data$w7==0] <- round(1/(1-predict(mod, type="response")[data$w7==0]),3)

summary(data$nr_weight_w7)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.138   1.248   1.376   2.001   1.651   7.911
```

```
mod <- glm(w8 ~ factor(agecat)+factor(gender)+factor(lregion)+factor(educ)
           +factor(partner)+factor(work), data=data, family=binomial())

data$nr_weight_w8 <- round(1/predict(mod, type="response"),3)
data$nr_weight_w8[data$w8==0] <- round(1/(1-predict(mod, type="response")[data$w8==0]),3)

summary(data$nr_weight_w8)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.142   1.287   1.371   2.002   2.308   7.060
```

```
mod <- glm(w9 ~ factor(agecat)+factor(gender)+factor(lregion)+factor(educ)
           +factor(partner)+factor(work), data=data, family=binomial())

data$nr_weight_w9 <- round(1/predict(mod, type="response"),3)
data$nr_weight_w9[data$w9==0] <- round(1/(1-predict(mod, type="response")[data$w9==0]),3)

summary(data$nr_weight_w9)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.098   1.210   1.368   2.006   1.826  11.168
```