

Appendix E: Weighting EVS17 data

Report related to the EVS 2017 integrated dataset ZA7500

- Current dataset version: v2-0-0, 2019-07-15 (second pre-release)
 - DOI: 10.4232/1.13314 (URL: <http://dx.doi.org/10.4232/1.13314>)
-

This document outlines the procedure used to compute the EVS17 weights and information on the source of the population statistics and adaptations required in each country.

Variable name and label:

gweight: 'Calibration weights'

pweight: 'Population size weights'

mm_pweight: 'Population size weights: mode'

ager_r_weight: 'CALIBRATION variable: Age in categories'

v225_weight: 'CALIBRATION variable: Gender'

v243_r_weight: 'CALIBRATION variable: Educational level'

NUTS_weight: 'CALIBRATION variable: Region at NUTS2 level'

Weighting EVS17 Data¹

Introduction

The weighting of survey data generally denotes every operation that alters the relative importance of single sampling units (or groups of units) for the purpose of estimating relevant statistics (e.g. means or totals) of a target population. The weights included in the second pre-release of EVS data are Calibration weights and population size weights. The type of analysis one is planning to conduct drives the choice of the weight, their combination is described more specifically in the table below.

| Research type | Weights to be used |
|--|---|
| Single country | Calibration weights |
| Country comparison with combined statistics | Calibration weights & Population size weights |
| Country comparison without combined statistics | Calibration weights |
| Country comparison with combined statistics and reference to groups of countries (i.e. region) | Calibration weights & Population size weights |

Calibration weights ('gweight')

Calibration weights serve various purposes. Two of the main reasons for that sort of weighting would be the potential reduction in an estimate's variance and the potential reduction in bias due to nonresponse/missing data. Hence, these weights often adjust (some) socio-demographic characteristics in the sample population to the distribution of the target population. These weights have been computed for each country on the basis of the population distribution provided by the countries themselves. The source of the population statistics are listed in the Appendix. Separated population margins of age, sex, educational levels and region have been employed. The description of the variables used is included in the Appendix. We computed this weighting variable employing the *calib* function of the R package *sampling* v2.8 by Alina Matei². The *Linear* method has been employed, which results also in negative weights³. Such a method requires the specification of each individual inclusion probability. Due to the lack of this information, we employed a proxy of the inclusion probability computed in each country, assuming the sample to be drawn randomly:

$$\text{Inclusion probability} = \frac{\text{Sample size}}{\text{Target Population size}}$$

The inverse of the resulting value has been employed as a vector of initial weights.

¹ The weighting procedure and the preparation of this document have been done in consultation with experts from GESIS – Leibniz Institute for the Social Sciences, in particular Matthias Sand and Christian Bruch and, in the early stages, Stefan Zins.

² <https://cran.r-project.org/web/packages/sampling/sampling.pdf>

³ That is a result of the statistical procedure, since a regression estimator is used to calibrate the survey data and the weights are then regression coefficients. We are, however, aware that unfortunately cases with negative weights are removed from the analyses by the most common statistical software such as SPSS or STATA. One option to get rid of negative weights, is to truncate the calibration weights to an upper and lower bound (e.g. 0.025- and 0.975-quantile). By truncating weights, one limits the range of the weights to a predetermined threshold while the sum of weights stays the same. However, this has then to be done for all countries (including those without negative weights) to assure comparability. Please be aware that this would introduce a larger bias into the estimation.

Further country specific deviations that occurred in the computation of weights are listed below in the Appendix. The dataset is provided including the variables with the imputed values for transparency and reproducibility purposes, a suffix “_weight” has been added to these variables.

Missing values imputation

The application of weighing procedures requires complete observations without missing values in the weighting variables. For this reason, missing values resulting from non-response on respondent’s gender, age and educational level, were imputed through the R package *mice*⁴. Our approach consisted in producing a single imputation based on 50 iterations, employing the *Classifications and regression trees* method⁵. The employment of multiple auxiliary variables is meant to get more accurate prediction of the imputed values on the basis of relevant respondent’s characteristics. The variables used as predictors for the imputation have been selected by the EVS central team because they represent core aspects of one’s value orientation (e.g. belonging to a religious denomination and interest in politics), or because they correlate with the weighting variables (e.g. Age at which full time education was completed and educational level). These variables are: *v51* (Belonging to a religious denomination), *v242* (Age at which full time education was completed), *v97* (Interest in politics), *v234* (Current marital status), *v227* (Nationality), *v225* (Respondent’s sex), *v243_r* (Respondent’s level of education in three categories) and *age_r3* (Respondent’s age in three categories). More details on these variables can be found in the ZA7500 Variable Report. In three countries (NL, IS and DE), in the CAWI-Mail mode, *v97* has been replaced by *v5* (Importance of politics in life), assuming that the two questions were capturing similar concepts, because of the matrix design.

Population size weights (‘pweight’ and ‘mm_pweight’)

Population size weights are provided for rescaling the weights to a shared denominator across all countries. These weights must be applied whenever one ought to analyze together different countries and avoids the overrepresentation of small countries when compared to bigger ones. The population size weight are computed as follows:

$$\text{Population size weight} = \frac{\text{Population size aged 18 years and older}}{\text{Net sample size} * 10\,000}$$

In this case the variables provided are two: the first variable “*pweight*” uses the full sample size of each country without distinguishing between modes. The second variable “*mm_pweight*” takes into account the mode structure, hence it splits the sample size conditionally on the mode for the six mixed-mode countries (CH, IS, NL, DE, FI, DK).

Table 4 in the Appendix lists the Population Size 18+ for each country, that can be used to compute additional versions of the population size weights for different subsamples.

⁴ <https://cran.r-project.org/web/packages/mice/mice.pdf>

⁵ More information in this regard can be found here: <https://www.jstatsoft.org/article/view/v045i03>.
EVS2017 Variable Report, App E (July 2019, v1)

Appendix

Calibration variables

The variables used for calibration are age, gender, region and level of education. The standard coding for those variables is described in the tables below. In some cases, the data provided at the country level did not match such coding. Consequently, the coding have been adjusted to match population data, the cases where this changes occurred are listed in the next section of the appendix.

| Code | agecat |
|------|--------|
| 1 | 18-24 |
| 2 | 25-34 |
| 3 | 35-44 |
| 4 | 45-54 |
| 5 | 55-64 |
| 6 | 65-74 |
| 7 | 75+ |

Table 1: Coding of age categories for calibration

| Code | Sex |
|------|--------|
| 1 | Male |
| 2 | Female |

Table 2. Coding of Gender for calibration

| Code | edu | Isced11 (or first digit ofedulvlb) | Description |
|------|--------|------------------------------------|--|
| 1 | Lower | 0, 1, 2 | Lower secondary or below |
| 2 | Medium | 3, 4 | Higher secondary and Post-secondary non-tertiary |
| 3 | Higher | 5, 6, 7, 8 | Post-secondary and above |

Table 3: Coding of educational levels for calibration

| Country code | Country | Target population size (source: population statistics) |
|---------------------|--------------------|--|
| AL | Albania | 2007877 |
| AM | Armenia | 2190686 |
| AT | Austria | 7509125 |
| AZ | Azerbaijan | 8550900 |
| BG | Bulgaria | 6181241 |
| BY | Belarus | 7304173 |
| CH | Switzerland | 6909664 |
| CZ | Czechia | 8585396 |
| DE | Germany | 68084270 |
| DK | Denmark | 4615690 |
| EE | Estonia | 1065731 |
| ES | Spain | 38310165 |
| FI | Finland | 4431392 |
| FR | France | 49764123 |
| GB | Great Britain | 50644094 |
| GE | Georgia | 5775216 |
| HR | Croatia | 3437119 |
| HU | Hungary | 7883893 |
| IS | Iceland | 241300 |
| IT | Italy | 50677616 |
| LT | Lithuania | 2595097 |
| NL | Netherlands | 13794988 |
| NO | Norway | 4183147 |
| PL | Poland | 31798656 |
| RO | Romania | 18267514 |
| RS | Serbia | 6017672 |
| RU | Russian Federation | 121153927 |
| SE | Sweden | 7998644 |
| SI | Slovenia | 1628479 |
| SK | Slovakia | 4432721 |

Table 4: Target population size by country

Country specific deviations in calibration variables

The section below lists the population data sources and all the country specific deviations both for individual and country level data. The algorithm employed for calculating the weights requires the total N of the marginal distributions of Age, Gender and Education to be equal. In those cases where the total N was different the margins have been corrected. The cases where this changes where applied are noted with “margins equalized”.

The item “missing values imputed” indicates how many observation had missing data on the variables of interest and have hence been imputed with the procedure described above. ‘None’ means there were no missing values on the variables of interest, hence there was no need for imputation.

Albania – AL

Source of data: N.A.

Population statistics: Standard classification, margins equalized; not weighted on regions due to lack of statistics at regional level.

Missing values imputed: 5 Age

Armenia – AM

Source of data: 2011 Armenian Population Census

Population statistics: Educational distribution starts at 20 years old, margins equalized.

Missing values imputed: None

Austria – AT

Source of data: 2015 Bildungsstandregister

Population statistics: Standard classification, margins equalized.

Missing values imputed: None

Azerbaijan – AZ

Source of data: Azərbaycan Respublikasının Dövlət Statistika Komitəsi (State Statistical Committee of the Republic of Azerbaijan)

Population statistics: Educational distribution ranges from 25 to 65 years old, different age categories, margins equalized.

Missing values imputed: None

Belarus – BY

Source of data: National Statistical Committee of the Republic of Belarus as of January 01, 2017 for age and gender. Survey data "Generations and Gender" for Education.

Population statistics: Standard classification.

Missing values imputed: 1 Education

Bulgaria – BG

Source of data: N.A.

Population statistics: Age categories stop at 60 years old.

Missing values imputed: 7 Education, 18 Age

Croatia – HR

Source of data: N.A.

Population statistics: ISCED code 2 coded as 'Medium', margins equalized.

Missing values imputed: 6 Education, 1 Age

Czech Republic – CZ

Source of data: Czech 2011 Population Census

Population statistics: Standard classification.

Missing values imputed: 66 Age, 17 Education

Denmark – DK

Source of data: Statistics Denmark

Population statistics: Standard classification, margins equalized

Missing values imputed: CAPI: 2 Age, 1 Sex, 12 Education | CAWI: 3 Age, 3 Sex, 67 Education

Estonia – EE

Source of data: Estonian Statistics.

Population statistics: Standard classification.

Missing values imputed: None

N.B. Please note that education data from Statistics Department is a mix of the 2011 census, population register (people may mention education voluntarily) and different education registers.

Finland – FI

Source of data: Tilastokeskus.

Population statistics: Standard classification; region FI20 missing in survey data hence dropped. Calibration weights computed on full CAPI+CAWI sample.

Missing values imputed: 2 Age, 2 Sex, 6 Education, 3 Region.

France – FR

Source of data: INSEE enquête Emploi 2014 – updated in 2017

Population statistics: Population distribution for education stops at 70 years old; not weighted on regions due to lack of statistics at regional level.

Missing values imputed: 5 Age, 7 Education

Georgia – GE

Source of data: N.A.

Population statistics: Standard classification, margins equalized.

Missing values imputed: 2 Education

Germany – DE

Source of data: German Microcensus 2016.

Population statistics: Standard classification, margins equalized. Regions are NUTS1.

Missing values imputed: CAPI: 1 Education | CAWI: 60 Age, 28 sex, 68 Education

Great Britain – GB

Source of data: N.A.

Population statistics: Standard classification, margins equalized.

Missing values imputed: 12 Age, 11 Education.

Hungary – HU

Source of data: N.A.

Population statistics: Refer to 20+ population, different age categories.

Missing values imputed: 8 Education

Iceland – IS

Source of data: N.A.

Population statistics: Education distribution ends at 74 years old.

Missing values imputed: CAPI: 5 Education | CAWI: 27 Education, 20 Sex, 22 Age

Italy – IT

Source of data: Eurostat (January 1, 2018.) for Age and Gender, ISTAT Census data 2011 for Education.

Population statistics: Standard classification, margins equalized. Regions are NUTS1.

Missing values imputed: 13 Education

Lithuania – LT

Source of data: N.A.

Population statistics: margins equalized; not weighted on regions due to lack of statistics at regional level.

Missing values imputed: 1 Age, 7 Education

Netherlands – NL

Source of data: *Centraal Bureau voor de Statistiek* for Age and Gender distribution, CBS statline for Education.

Population statistics: Standard classification, margins equalized. CAWI sample not weighted on region because of missing information in the data.

Missing values imputed: CAPI: 3 Education, CAWI: 43 Edu, 1 Age, 1 Sex

Norway – NO

Source of data: Statistics Norway (Education: 2018; age and gender: NA)

Population statistics: margins equalized; different age categories, not weighted on regions due to lack of statistics at regional level.

Missing values imputed: 4 Age, 22 Education.

Poland – PL

Source of data: PESEL (Universal Electronic System for Registration of the Population) registry for Age and Gender, Labour force survey in Poland I quarter 2018 for Education.

Population statistics: Standard classification, margins equalized.
Missing values imputed: 8 Education

Romania – RO

Source of data: Institutul național de statistică; Education from 2011 Census
Population statistics: Standard classification, margins equalized.
Missing values imputed: 40 Education, 54 Age

Russia – RU

Source of data: Russian National Census 2010.
Population statistics: Standard classification, margins equalized.
Missing values imputed: 7 Education

Serbia – RS

Source of data: N.A.
Population statistics: Standard classification, margins equalized.
Missing values imputed: 19 Age, 8 Education

Slovakia – SK

Source of data: N.A.
Population statistics: Age classification ends at 65 years old.
Missing values imputed: 10 Education

Slovenia – SI

Source of data: Statistical Office of the Republic of Slovenia
Population statistics: Age starts at 20 and ends at 65, margins equalized.
Missing values imputed: 7 Education

Spain – ES

Source of data: Encuesta del Padrón, INE; Encuesta de Población Activa, INE
Population statistics: Population distribution for education starts at 25 years old; margins equalized.
Missing values imputed: 7 Education

Sweden – SE

Source of data: N.A.
Population statistics: Standard classification, margins equalized.
Missing values imputed: 14 Education, 2 Sex, 8 Age

Switzerland – CH

Source of data: Statistik der Bevölkerung und der Haushalte (STATPOP).
Population statistics: ISCED codes 3, 4, 5 aggregated in the 'Medium' category. Calibration weights computed on full CAPI+CAWI sample.
Missing values imputed: 2 Education, 2 Age

Example of R script to compute weights

```
##### Weights: Master Syntax #####

if (!require(sampling)) install.packages('sampling')
if (!require(dplyr)) install.packages('dplyr')
if (!require(haven)) install.packages('haven')
if (!require(readxl)) install.packages('readxl')
if (!require(magrittr)) install.packages('magrittr')
if (!require(mice)) install.packages('mice')

# Replace working directory
setwd("directory") # !!!! remember to use / and not \

# Load packages -----
library(sampling)
library(dplyr)
library(haven)
library(readxl)
library(magrittr)
library(mice)

# Load datasets -----
# Load Integrated Data Set EVS2017 (IDS) (contains 16 first-release countries)
ids <- read_sav("ZA 7500/ZA7500_v1-0-0.sav")

# Select the country by replacing the number
sam <- ids[ which(ids$country==000), ]

# Load Excel file according to information provided
# 1. Syntax for ASE
read_excel("*.xlsx",
           sheet = "ASE_by_R") %>%
  mutate(RxAxSxE=paste(region, agecat, sex, edu, sep="."),
         A_Freq=tapply(N_group,agecat,sum)[agecat],
         S_Freq=tapply(N_group,sex,sum)[sex],
         E_Freq=tapply(N_group,edu,sum)[edu],
         R_Freq=tapply(N_group,region,sum)[region]) %>%
  rename(RxAxSxE_Freq=N_group) ->
  pop_ASEbyR

# 2. Syntax for AS + E
read_excel("*.xlsx",
           sheet = "AS_by_R") %>%
  mutate(RxAxS=paste(region, agecat, sex, sep="."),
         A_Freq=tapply(N_group,agecat,sum)[agecat],
         S_Freq=tapply(N_group,sex,sum)[sex],
         R_Freq=tapply(N_group,region,sum)[region]) %>%
  rename(RxAxS_Freq=N_group) ->
  pop_ASbyR

read_excel("*.xlsx",
           sheet = "E_by_R") %>%
  mutate(RxE=paste(region,edu,sep="."),
         E_Freq=tapply(N_group,edu,sum)[edu]) %>%
  rename(RxE_Freq=N_group) ->
  pop_EbyR

# 3. Syntax for A + S + E by Region
read_excel("*.xlsx",
           sheet = "A_by_R") %>%
  mutate(A=paste(region, agecat, sep="."),
         A_Freq=tapply(N_group,agecat,sum)[agecat],
         R_Freq=tapply(N_group,region,sum)[region]) %>%
  rename(A_Freq=N_group) ->
  pop_AbyR
```

```

read_excel("****.xlsx",
            sheet = "S_by_R") %>%
  mutate(Rxs=paste(region, sex, sep="."),
         S_Freq=tapply(N_group,sex,sum)[sex],
         R_Freq=tapply(N_group,region,sum)[region]) %>%
  rename(S_Freq=N_group) ->
  pop_SbyR

read_excel("****.xlsx",
            sheet = "E_by_R") %>%
  mutate(E=paste(region,edu,sep="."),
         E_Freq=tapply(N_group,edu,sum)[edu],
         R_Freq=tapply(N_group,region,sum)[region]) %>%
  rename(E_Freq=N_group) ->
  pop_EbyR

# 4. Syntax for A + S + E + R separated
read_excel("****.xlsx",
            sheet = "A") %>%
  mutate(A=paste(region, agecat, sep="."),
         A_Freq=tapply(N_group,agecat,sum)[agecat]) %>%
  rename(A_Freq=N_group) ->
  pop_A

read_excel("****.xlsx",
            sheet = "S") %>%
  mutate(Rxs=paste(region, sex, sep="."),
         S_Freq=tapply(N_group,sex,sum)[sex]) %>%
  rename(S_Freq=N_group) ->
  pop_S

read_excel("****.xlsx",
            sheet = "E") %>%
  mutate(E=paste(region,edu,sep="."),
         E_Freq=tapply(N_group,edu,sum)[edu]) %>%
  rename(E_Freq=N_group) ->
  pop_E

read_excel("****.xlsx",
            sheet = "R") %>%
  mutate(R=paste(region,sep="."),
         R_Freq=tapply(N_group,region,sum)[region]) %>%
  rename(R_Freq=N_group) ->
  pop_R

# Prepare sample data
sam %>%
  select(id_cocas, v225, age_r3, v243_r, v275b_N2) %>%
  mutate(v225 = factor(v225),
         age_r3 = factor(age_r3),
         v243_r = factor(v243_r),
         v275b_N2= factor(v275b_N2) ) %>%
  rename(agecat=age_r3,
         sex=v225,
         edu=v243_r,
         region=v275b_N2) -> sam_net

# Check for missing values
count(sam_net, agecat)
count(sam_net, sex)
count(sam_net, edu)
count(sam_net, region)

# Imputation routine -----
# Select variables

```

```

sam %>%
  select(caseno, country, id_cocas, v51, v242, v97, v234, v227, v225, age_r3, v243_r, v275b_N2) %>%
  mutate(v225 = factor(v225),
         age_r3 = factor(age_r3),
         v243_r = factor(v243_r),
         v275b_N2= factor(v275b_N2),
         v51 = factor(v51),
         v242 = factor(v242),
         v97 = factor(v97),
         v234 = factor(v234),
         v227 = factor(v227)) %>%
  rename(agecat=age_r3,
         sex=v225,
         edu=v243_r,
         region=v275b_N2) -> imp_data

count(imp_data, is.na(v51))
count(imp_data, is.na(v242))
count(imp_data, is.na(v97))
count(imp_data, is.na(v234))
count(imp_data, is.na(v227))

# First run mice (with only 1 iteration) to create the predictor matrix
sam_imp <- mice(imp_data,m=1,maxit=1,meth='cart',seed=500)
summary(sam_imp)

# Set ids and country to 0 (they do not impute)
pred_matrix <-sam_imp$predictorMatrix
pred_matrix[, 1] <- 0

# Leave only imputation of agecat and edu (and sex and region)
pred_matrix[1:8,] <- 0

# Run mice again using new predictor matrix and the full number of iterations
sam_imp_reduced <- mice(sam_imp$data, m=1, maxit=50, pred_matrix, meth='cart',seed=500)

# Check imputations
sam_imp_reduced$imp$agecat
sam_imp_reduced$imp$edu
sam_imp_reduced$imp$sex

cbind(as.data.frame(sam_imp_reduced$imp$edu),
      subset(imp_data, is.na(edu))) -> check_imp_edu

cbind(as.data.frame(sam_imp_reduced$imp$agecat),
      subset(imp_data, is.na(agecat))) -> check_imp_agecat

cbind(as.data.frame(sam_imp_reduced$imp$sex),
      subset(imp_data, is.na(sex))) -> check_imp_sex

check_imp_edu
check_imp_agecat
check_imp_sex

# Adding imputed values to dataframe
sam_comp <- mice::complete(sam_imp_reduced, 1)

# Selecting only relevant columns
sam_comp %>%
  select(caseno, country, id_cocas, sex, agecat, edu, region) -> sam_comp

count(sam_comp, agecat)
count(sam_comp, sex)
count(sam_comp, edu)
count(sam_comp, region)

# Dropping useless data
rm(check_imp_edu, check_imp_agecat, sam_imp,sam_imp_reduced, imp_data)

```

```

# Prepare population margins -----
# prepare sample data for weighting procedure

sam_comp$agecat <- droplevels(sam_comp $agecat)
sam_comp$edu <- droplevels(sam_comp$edu)
sam_comp$sex <- droplevels(sam_comp$sex)

Xs <-
  model.matrix(~ sex+agecat+edu+region-1,
              data=sam_comp,
              contrasts.arg=list(sex=contrasts(sam_comp$sex, contrasts=F),
                                agecat=contrasts(sam_comp$agecat, contrasts=F),
                                edu=contrasts(sam_comp$edu, contrasts=F),
                                region=contrasts(sam_comp$region, contrasts=F))
              )

# Aggregating Population Margins following the procedure above
# 1. ASE
mar_ASER <-
  with(pop_ASEbyR,
       c(tapply(RxAxSxE_Freq,
                paste0("sex", sex),
                sum),
         tapply(RxAxSxE_Freq,
                paste0("agecat", agecat),
                sum),
         tapply(RxAxSxE_Freq,
                paste0("edu", edu),
                sum),
         tapply(RxAxSxE_Freq,
                paste0("region", region),
                sum)
       ) )

# 2. AS + E
mar_AS <-
  with(pop_ASbyR,
       c(tapply(RxAxS_Freq,
                paste0("sex", sex),
                sum),
         tapply(RxAxS_Freq,
                paste0("agecat", agecat),
                sum)
       ) )
mar_E <-
  with(pop_EbyR,
       c(tapply(E_Freq,
                paste0("edu", edu),
                sum)
       ) )

# 3. A + S + E
mar_S <-
  with(pop_SbyR,
       c(tapply(RxS_Freq,
                paste0("sex", sex),
                sum)
       ) )

mar_A <-
  with(pop_AbyR,
       c(tapply(RxA_Freq,
                paste0("agecat", agecat),
                sum) ) )

mar_E <-
  with(pop_EbyR,
       c(tapply(RxE_Freq,
                paste0("edu", edu),

```

```

        sum)
    ) )

# Are population margins equal?

# 2: AS + E
mar_ASR
sum(mar_ASR[1:2])
sum(mar_E)

# 3: A + S + E
mar_E
mar_A
mar_S
mar_R

# If no:
a <- sum('right'[1:2])
b <- sum('wrong')

'wrong' <- ('wrong'/ b)* a

# Binding Population Margins together
# 1. ASE
mar_pop <- (mar_ASER)
# 2. AS + E
mar_pop <- c(mar_AS, mar_ER)
# 3. A + S + E
mar_pop <- c(mar_S, mar_A, mar_E, mar_R)

# Computing weights -----
# Total population
totpop <- sum(mar_pop[1:2])

# Sample population
n <- nrow(Xs)

# Average design weight
p <- totpop/n

# Computing calibration weights
gweight <- calib(as.matrix(Xs),
                d = rep(p,nrow(Xs) ),
                mar_pop,
                method = "linear")

# Check weights
checkcalibration(Xs,d=rep(p,nrow(Xs)),mar_pop,gweight)$result

summary(gweight)

# Population size weight -----
totpop
pweight <- totpop/(n*10000) # if not mixed mode
# mm_pweight <- totpop/(n*10000) # if mixed mode
# n_tot <- XXXX # insert full sample size - if mixed mode
# pweight <- totpop/(n_tot*10000) # if mixed mode

# Export dataset in SPSS format -----
sam_comp$gweight <- gweight

# ! If not mixed mode:
sam_comp$pweight <- pweight
sam_comp$mm_pweight <- pweight

# ! If mixed mode:
# sam_comp$pweight <- pweight
# sam_comp$mm_pweight <- mm_pweight

```

```
sam_comp$NUTS_weight <-rep(NA, ****)
sam_comp$NUTS_weight <-as.character((sam$v275b_N2))
sam_comp$ager_r_weight <- sam_comp$agecat
sam_comp$v225_weight <-sam_comp$sex
sam_comp$v243_r_weight <-sam_comp$edu

sam_comp <- subset(sam_comp, select = -c(sex, edu, agecat, region))

glimpse(sam_comp)

write_sav(sam_comp, "Weight datasets/weight_CC.sav") # put country name

# END
```