UKHLS weighted kin 4.6

Reading and cleaning data

```
#houskeeping
#Clear objects already in the environment - start with a clean slate
rm(list=ls())

#loading libraries
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.3     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to beco
```

```
library(svyVGAM)
```

```
Loading required package: VGAM
Loading required package: stats4
Loading required package: splines
Loading required package: survey
Loading required package: grid
Loading required package: Matrix
```

```
Attaching package: 'Matrix'

The following objects are masked from 'package:tidyr':

    expand, pack, unpack

Loading required package: survival

Attaching package: 'survey'

The following object is masked from 'package:VGAM':

    calibrate

The following object is masked from 'package:graphics':

    dotchart
```

```
library(sjlabelled)
```

```
Warning: package 'sjlabelled' was built under R version 4.3.3


Attaching package: 'sjlabelled'

The following object is masked from 'package:forcats':

    as_factor

The following object is masked from 'package:dplyr':

    as_label

The following object is masked from 'package:ggplot2':

    as_label
```

```
library(desctable)
```

```
Warning: package 'desctable' was built under R version 4.3.3
```

```
Loading required package: pander
```

```
Warning: package 'pander' was built under R version 4.3.3
```

```

Attaching package: 'desctable'
```

```
The following objects are masked from 'package:stats':
```

```
    chisq.test, fisher.test, IQR
```

```
library(summarytools)
```

```
Warning: package 'summarytools' was built under R version 4.3.3
```

```

Attaching package: 'summarytools'
```

```
The following object is masked from 'package:sjlabelled':
```

```
    unlabel
```

```
The following object is masked from 'package:tibble':
```

```
    view
```

```
library(naniar)
```

```
Warning: package 'naniar' was built under R version 4.3.3
```

```
library(survey)
library(svrep)
library(Hmisc)
```

```
Attaching package: 'Hmisc'

The following objects are masked from 'package:summarytools':

    label, label<-

The following object is masked from 'package:survey':

    deff

The following objects are masked from 'package:dplyr':

    src, summarize

The following objects are masked from 'package:base':

    format.pval, units
```

```
library(srvyr)
```

```
Attaching package: 'srvyr'

The following object is masked from 'package:Hmisc':

    summarize

The following object is masked from 'package:stats':

    filter
```

```
library(marginaleffects)
library(haven)
```

```
Attaching package: 'haven'

The following objects are masked from 'package:sjlabelled':

    as_factor, read_sas, read_spss, read_stata, write_sas, zap_labels
```

```r
library(catregs)
library(margins)
library(modelsummary)
```

```
Attaching package: 'modelsummary'

The following object is masked from 'package:Hmisc':

    Mean

The following object is masked from 'package:VGAM':

    Max
```

```r
library(zoo)
```

```
Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
```

```r
library(mice)
```

```
Warning: package 'mice' was built under R version 4.3.3

Attaching package: 'mice'

The following object is masked from 'package:stats':

    filter

The following objects are masked from 'package:base':

    cbind, rbind
```

```
library(stargazer)
```

```
Please cite as:

 Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
 R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(texreg)
```

```
Warning: package 'texreg' was built under R version 4.3.3
```

```
Version:  1.39.3
Date:     2023-11-09
Author:   Philip Leifeld (University of Essex)

Consider submitting praise using the praise or praise_interactive functions.
Please cite the JSS article in your publications -- see citation("texreg").

Attaching package: 'texreg'

The following object is masked from 'package:tidyr':

    extract
```

```
library(VIM)
```

```
Warning: package 'VIM' was built under R version 4.3.3
```

```
Loading required package: colorspace
VIM is ready to use.

Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues

Attaching package: 'VIM'

The following object is masked from 'package:VGAM':
```

```
    wine

The following object is masked from 'package:datasets':

    sleep
```

```
library(lattice)
library(ggplot2)
```

Cross-wave dat

```
inpath<-"D:/r git projects/ox-R/final essay/UKHLS n BHPS stata/UKDA-6614-stata/stata/stata
xwavedat<-read_dta(file=paste0(inpath, "ukhls/xwavedat.dta")) %>%
 dplyr::select(pidp,maju,paju,maedqf,paedqf)
```

```
missval <- c(-9, -8, -7, -2, -1)
for (i in 1:5) {
  xwavedat<- xwavedat %>%
    mutate_all(., list(~na_if(., missval[i])))

}
```

Wave1

```
#wave 1
a_indresp <- read_dta(file=paste0(inpath, "ukhls/a_indresp.dta")) %>%
 dplyr::select(pidp, a_mastat_dv, a_strata, a_psu,
         a_indinus_xw,
         a_age_dv, a_sex_dv, a_hiqual_dv,
         a_mastat_dv, a_racel_dv,a_fimnnet_dv,a_lvrel1,a_lvrel2,a_pns1pno,a_pns2pno,a_nchi


a_indrespj<-a_indresp%>%dplyr::select(pidp,a_panssec8_dv,a_manssec8_dv)

library(plyr)
```

```
You have loaded plyr after dplyr - this is likely to cause problems.
If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
library(plyr); library(dplyr)

------------------------------------------------------------------------------


Attaching package: 'plyr'

The following objects are masked from 'package:srvyr':

    mutate, rename, summarise, summarize

The following objects are masked from 'package:Hmisc':

    is.discrete, summarize

The following objects are masked from 'package:dplyr':

    arrange, count, desc, failwith, id, mutate, rename, summarise,
    summarize

The following object is masked from 'package:purrr':

    compact
```

```
a_indresp<-join_all(list(a_indresp,xwavedat), by='pidp', type='left')
```

```
#rename pa ses, job and edu columns
a_indresp <-a_indresp%>%
  dplyr::rename(
   pases=a_panssec8_dv,
    mases=a_manssec8_dv
    )
```

Wave 2~12

```r
#wave 2
b_indresp <- read_dta(file=paste0(inpath, "ukhls/b_indresp.dta")) %>%
 dplyr::select(pidp,b_panssec8_dv,b_manssec8_dv)

#wave 3
c_indresp <- read_dta(file=paste0(inpath, "ukhls/c_indresp.dta")) %>%
 dplyr::select(pidp,c_panssec8_dv,c_manssec8_dv)

#wave 4
d_indresp <- read_dta(file=paste0(inpath, "ukhls/d_indresp.dta")) %>%
 dplyr::select(pidp,d_panssec8_dv,d_manssec8_dv)

#wave 5
e_indresp <- read_dta(file=paste0(inpath, "ukhls/e_indresp.dta")) %>%
 dplyr::select(pidp,e_panssec8_dv,e_manssec8_dv)

#wave 6
f_indresp <- read_dta(file=paste0(inpath, "ukhls/f_indresp.dta")) %>%
 dplyr::select(pidp,f_panssec8_dv,f_manssec8_dv)

#wave 7
g_indresp <- read_dta(file=paste0(inpath, "ukhls/g_indresp.dta")) %>%
 dplyr::select(pidp,g_panssec8_dv,g_manssec8_dv)

#wave 8
h_indresp <- read_dta(file=paste0(inpath, "ukhls/h_indresp.dta")) %>%
 dplyr::select(pidp,h_panssec8_dv,h_manssec8_dv)

#wave 9
i_indresp <- read_dta(file=paste0(inpath, "ukhls/i_indresp.dta")) %>%
 dplyr::select(pidp,i_panssec8_dv,i_manssec8_dv)

#wave 10
j_indresp <- read_dta(file=paste0(inpath, "ukhls/j_indresp.dta")) %>%
 dplyr::select(pidp,j_panssec8_dv,j_manssec8_dv)

#wave 11
k_indresp <- read_dta(file=paste0(inpath, "ukhls/k_indresp.dta")) %>%
 dplyr::select(pidp,k_panssec8_dv,k_manssec8_dv)

#wave 12
```

```
l_indresp <- read_dta(file=paste0(inpath, "ukhls/l_indresp.dta")) %>%
 dplyr::select(pidp,l_panssec8_dv,l_manssec8_dv)
```

Wave13

```
#wave 13
m_indresp <- read_dta(file=paste0(inpath, "ukhls/m_indresp.dta")) %>%
 dplyr::select(pidp, m_mastat_dv, m_strata, m_psu,
         m_indpxui_xw,
         m_age_dv, m_sex_dv, m_hiqual_dv,
         m_mastat_dv, m_racel_dv,m_fimnnet_dv,m_lvrel1,m_lvrel2,m_lvrel9,m_lvrel10,m_pns1p
```

```
# recode missing values for all variables using for loop and mutate_all
for (i in 1:5) {
  a_indresp<- a_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}

for (i in 1:5) {
  a_indrespj<- a_indrespj %>%
    mutate_all(., list(~na_if(., missval[i])))
}

for (i in 1:5) {
  b_indresp<- b_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  c_indresp<- c_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  d_indresp<- d_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  e_indresp<- e_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
```

```
  f_indresp<- f_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  g_indresp<- g_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  h_indresp<- h_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  i_indresp<- i_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  j_indresp<- j_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  k_indresp<- k_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  l_indresp<- l_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))
}
for (i in 1:5) {
  m_indresp<- m_indresp %>%
    mutate_all(., list(~na_if(., missval[i])))

}
```

```
#wave 13 join
library(plyr)
m_indresp<-join_all(list(m_indresp,a_indrespj,b_indresp,c_indresp,d_indresp,e_indresp,f_in

m_indresp<-m_indresp%>% mutate(
                            mases=coalesce(a_manssec8_dv,b_manssec8_dv,c_manssec8_dv,d_mans
                            pases=coalesce(a_panssec8_dv,b_panssec8_dv,c_panssec8_dv,d_pans
```

```
Warning: `..1` and `..2` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..3` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..4` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..5` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..6` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..7` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..8` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..9` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..10` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..11` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7
```

```
Warning: `..1` and `..12` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7

Warning: `..1` and `..13` have conflicting value labels.
i Labels for these values will be taken from `..1`.
x Values: -7
```

```
rm(missval,a_indrespj,b_indresp,c_indresp,d_indresp,e_indresp,f_indresp,g_indresp,h_indres
```

Cleaning data

Wave 1

```
a_indresp$a_sex_dv[a_indresp$a_sex_dv==0]<-NA
a_indresp$a_fimnnet_dv[a_indresp$a_fimnnet_dv<0]<-NA
```

```
#preparing variables for analysis (wave 1)
#num of parents
a_indresp$lvpam<-c(0)
a_indresp$lvpam[a_indresp$a_pns1pno!=0]<-1
a_indresp$lvpaf<-c(0)
a_indresp$lvpaf[a_indresp$a_pns2pno!=0]<-1
a_indresp$parentnum<-a_indresp$a_lvrel1+a_indresp$a_lvrel2+a_indresp$lvpam+a_indresp$lvpaf


#living with parents or not (biological/step/adoptive)
a_indresp$lvpa<-a_indresp$lvpam+a_indresp$lvpaf
a_indresp<-a_indresp%>%dplyr::select(-lvpam,-lvpaf)

#binary
a_indresp$lvpab<-c(0)
a_indresp$lvpab[a_indresp$lvpa!=0]<-1

#level of education completed
a_indresp$edu<-c(0)
a_indresp$edu[a_indresp$a_hiqual_dv%in%c(9)]<-1 #low (everything below)
a_indresp$edu[a_indresp$a_hiqual_dv%in%c(3,4,5)]<-2 #middle (completed A-level or secondar
a_indresp$edu[a_indresp$a_hiqual_dv%in%c(1,2)]<-3 #high (have a degree)
```

```r
a_indresp$edu[a_indresp$edu==0]<-NA

print(attr(a_indresp$a_hiqual_dv,"labels"))
```

```
    missing inapplicable      refusal   don't know       Degree Other higher
         -9           -8           -2           -1            1            2
A level etc     GCSE etc   Other qual      No qual
          3            4            5            9
```

```r
a_indresp$edu<-factor(a_indresp$edu,levels=c(1,2,3),labels=c('low',"middle","high"))


#making nchild_dv a binary variable
a_indresp$child<-c(0)
a_indresp$child[a_indresp$a_nchild_dv>=1]<-1

#age group
a_indresp$age_group<-c(0)
a_indresp$age_group[a_indresp$a_age_dv%in%c(16:20)]<-1
a_indresp$age_group[a_indresp$a_age_dv%in%c(21:25)]<-2
a_indresp$age_group[a_indresp$a_age_dv%in%c(26:30)]<-3
a_indresp$age_group[a_indresp$a_age_dv%in%c(31:35)]<-4

a_indresp$age_group<-factor(a_indresp$age_group,levels=c(1,2,3,4),labels=c("16-20","21-25"

#logarithm of income
a_indresp$lnincome<-log(a_indresp$a_fimnnet_dv)
#add 1 to all income values to avoid -inf in log transformation
a_indresp$incomeadd<-a_indresp$a_fimnnet_dv+1
a_indresp$lnincome<-log(a_indresp$incomeadd)
a_indresp<-a_indresp%>%dplyr::select(-incomeadd)

#racel groups
a_indresp$racel<-c(0)
a_indresp$racel[a_indresp$a_racel_dv%in%c(1,2,3,4)]<-1 #Whites
a_indresp$racel[a_indresp$a_racel_dv%in%c(14,15)]<-2 #African Caribbean
a_indresp$racel[a_indresp$a_racel_dv%in%c(9,10,11)]<-3 #Indiani,Pakistani, Bangladeshi
a_indresp$racel[a_indresp$a_racel_dv%in%c(12,13)]<-4 #Other Asian
a_indresp$racel[a_indresp$racel==0]<-5 #Other

a_indresp$racel<-factor(a_indresp$racel,levels=c(1,2,3,4,5),labels=c("Whites","African Car
```

```
#father ses
a_indresp$fases<-c(0)
a_indresp$fases[a_indresp$pases==8]<-1 #Routine
a_indresp$fases[a_indresp$pases==7]<-2 #Semi-routine
a_indresp$fases[a_indresp$pases==6]<-3 #Lower supervisory & technical
a_indresp$fases[a_indresp$pases==5]<-4 #Small employers & own account
a_indresp$fases[a_indresp$pases==4]<-5 #Intermediate
a_indresp$fases[a_indresp$pases==3]<-6 #Lower management & professional
a_indresp$fases[a_indresp$pases==2]<-7 #Higher professional
a_indresp$fases[a_indresp$pases==1]<-8 #Large employers & higher management

#dummy variable for pases missing
#a_indresp$fasesna<-ifelse(a_indresp$fases==0, 1, 0)
a_indresp$fases[a_indresp$pases==0]<-NA

a_indresp$fasesXage<-a_indresp$fases*a_indresp$a_age_dv

#for parental education
#father
a_indresp$paedu<-c(0)
a_indresp$paedu[a_indresp$paedqf%in%c(1,2,97)]<-1 #low (everything below, the category "ot
a_indresp$paedu[a_indresp$paedqf%in%c(3,4)]<-2 #middle (complete A-level or have some qual
a_indresp$paedu[a_indresp$paedqf%in%c(5)]<-3 #high (have a degree)
a_indresp$paedu[a_indresp$paedu==0]<-NA

a_indresp$paedu<-factor(a_indresp$paedu,levels=c(1,2,3),labels=c('low',"middle","high"))

#mother
a_indresp$maedu<-c(0)
a_indresp$maedu[a_indresp$maedqf%in%c(1,2,97)]<-1 #low (everything below, the category "ot
a_indresp$maedu[a_indresp$maedqf%in%c(3,4)]<-2 #middle (complete A-level or have some qual
a_indresp$maedu[a_indresp$maedqf%in%c(5)]<-3 #high (have a degree)
a_indresp$maedu[a_indresp$maedu==0]<-NA

a_indresp$maedu<-factor(a_indresp$maedu,levels=c(1,2,3),labels=c('low',"middle","high"))

#for whether mother working at age 14
a_indresp$mawork<-c(0)
a_indresp$mawork[a_indresp$maju==2]<-1 #notworking
a_indresp$mawork[a_indresp$maju==1]<-2 #working
a_indresp$mawork[a_indresp$maju%in%c(3,4)]<-3 #Mother deceased/Mother not living with resp
```

```r
a_indresp$mawork[a_indresp$mawork==0]<-NA #missing

a_indresp$mawork<-factor(a_indresp$mawork,levels=c(1,2,3),labels=c("notworking","working",
```

```r
#preparing dependent variables for multinomial regression (wave 1)
##patterns of home-leaving
###living with parents(including those who have children)=======================
a_indresp$leave[a_indresp$lvpa!=0]<-1
##route out 1: to live with a partner ============================
a_indresp$leave[a_indresp$lvpa==0 & #not living with parents
                        a_indresp$a_mastat_dv%in%c(2,3,10)  #married
                        ]<-2

##route out 2: to live without a partner=====================
a_indresp$leave[a_indresp$lvpa==0 & #not living with parents
                        a_indresp$a_mastat_dv%in%c(0,1,4,5,6,7,8,9,10)
                        ]<-3
```

```r
#preparing dependent variables for multinomial regression (wave 1)
##patterns of home-leaving (distinguish between marriage and cohabitation)
a_indresp$ledes<-c(0)

###living with parents(including those who have children)=======================
a_indresp$ledes[a_indresp$lvpa!=0]<-1

##route out 1: to live with a partner(marriage) ============================
a_indresp$ledes[a_indresp$lvpa==0 & #not living with parents
                        a_indresp$a_mastat_dv%in%c(2,3)  #married
                        ]<-2

##route out 2: to live with a partner(cohabitation) ============================
a_indresp$ledes[a_indresp$lvpa==0 & #not living with parents
                        a_indresp$a_mastat_dv%in%c(10)  #in cohabitation
                        ]<-3

##route out 3: to live without a partner====================
a_indresp$ledes[a_indresp$lvpa==0 & #not living with parents
                        a_indresp$a_mastat_dv%in%c(0,1,4,5,6,7,8,9)  #others
                        ]<-4
```

```
attr(a_indresp$ledes,"labels")<-c(
"NA:n=13,having missing values in marital status;",
"1:living with parents;",
"2:route out 1:leaving with partner (marriage) ;",
"3:route out 2:leaving with partner (cohabitation);",
"4:route out 3:leaving without a partner")


a_indresp$ledes<-factor(a_indresp$ledes,levels=c(1,2,3,4),labels=c("living with parents","
```

Wave 13

```
m_indresp<-m_indresp%>%dplyr::select(pidp, m_mastat_dv, m_strata, m_psu,
         m_indpxui_xw,
         m_age_dv, m_sex_dv, m_hiqual_dv,
         m_mastat_dv, m_racel_dv,m_fimnnet_dv,m_lvrel1,m_lvrel2,m_lvrel9,m_lvrel10,m_pns1p
```

```
m_indresp$m_sex_dv[m_indresp$m_sex_dv==0]<-NA
m_indresp$m_fimnnet_dv[m_indresp$m_fimnnet_dv<0]<-NA
```

```
#preparing variables for analysis (wave 13)
#num of parents
m_indresp$lvpam<-c(0)
m_indresp$lvpam[m_indresp$m_pns1pno!=0]<-1
m_indresp$lvpaf<-c(0)
m_indresp$lvpaf[m_indresp$m_pns2pno!=0]<-1
m_indresp$parentnum<-m_indresp$m_lvrel1+m_indresp$m_lvrel2+m_indresp$lvpam+m_indresp$lvpaf


#living with parents or not (biological/step/adoptive)
m_indresp$lvpa<-m_indresp$lvpam+m_indresp$lvpaf
m_indresp<-m_indresp%>%dplyr::select(-lvpam,-lvpaf)

#binary
m_indresp$lvpab<-c(0)
m_indresp$lvpab[m_indresp$lvpa!=0]<-1

#level of education completed
```

```r
m_indresp$edu<-c(0)
m_indresp$edu[m_indresp$m_hiqual_dv%in%c(9)]<-1 #low (everything below)
m_indresp$edu[m_indresp$m_hiqual_dv%in%c(3,4,5)]<-2 #middle (completed A-level or secondar
m_indresp$edu[m_indresp$m_hiqual_dv%in%c(1,2)]<-3 #high (have a degree)
m_indresp$edu[m_indresp$edu==0]<-NA

print(attr(m_indresp$m_hiqual_dv,"labels"))
```

```
           missing        inapplicable            refusal          don't know
                -9                  -8                  -2                  -1
            Degree Other higher degree         A-level etc           GCSE etc
                 1                   2                   3                   4
Other qualification    No qualification
                 5                   9
```

```r
m_indresp$edu<-factor(m_indresp$edu,levels=c(1,2,3),labels=c('low',"middle","high"))


#making nchild_dv a binary variable
m_indresp$child<-c(0)
m_indresp$child[m_indresp$m_nchild_dv>=1]<-1

#age group
m_indresp$age_group<-c(0)
m_indresp$age_group[m_indresp$m_age_dv%in%c(16:20)]<-1
m_indresp$age_group[m_indresp$m_age_dv%in%c(21:25)]<-2
m_indresp$age_group[m_indresp$m_age_dv%in%c(26:30)]<-3
m_indresp$age_group[m_indresp$m_age_dv%in%c(31:35)]<-4

m_indresp$age_group<-factor(m_indresp$age_group,levels=c(1,2,3,4),labels=c("16-20","21-25"

#logarithm of income
m_indresp$lnincome<-log(m_indresp$m_fimnnet_dv)
#add 1 to all income values to avoid -inf in log transformation
m_indresp$incomeadd<-m_indresp$m_fimnnet_dv+1
m_indresp$lnincome<-log(m_indresp$incomeadd)
m_indresp<-m_indresp%>%dplyr::select(-incomeadd)

#racel groups
m_indresp$racel<-c(0)
m_indresp$racel[m_indresp$m_racel_dv%in%c(1,2,3,4)]<-1 #Whites
```

```
m_indresp$racel[m_indresp$m_racel_dv%in%c(14,15)]<-2 #African Caribbean
m_indresp$racel[m_indresp$m_racel_dv%in%c(9,10,11)]<-3 #Indiani,Pakistani, Bangladeshi
m_indresp$racel[m_indresp$m_racel_dv%in%c(12,13)]<-4 #Other Asian
m_indresp$racel[m_indresp$racel==0]<-5 #Other

m_indresp$racel<-factor(m_indresp$racel,levels=c(1,2,3,4,5),labels=c("Whites","African Car

#
```

```
#father ses
m_indresp$fases<-c(0)
m_indresp$fases[m_indresp$pases==8]<-1 #Routine
m_indresp$fases[m_indresp$pases==7]<-2 #Semi-routine
m_indresp$fases[m_indresp$pases==6]<-3 #Lower supervisory & technical
m_indresp$fases[m_indresp$pases==5]<-4 #Small employers & own account
m_indresp$fases[m_indresp$pases==4]<-5 #Intermediate
m_indresp$fases[m_indresp$pases==3]<-6 #Lower management & professional
m_indresp$fases[m_indresp$pases==2]<-7 #Higher professional
m_indresp$fases[m_indresp$pases==1]<-8 #Large employers & higher management

#dummy variable for pases missing
#m_indresp$fasesna<-ifelse(m_indresp$fases==0, 1, 0)
m_indresp$fases[m_indresp$pases==0]<-NA

m_indresp$fasesXage<-m_indresp$fases*m_indresp$m_age_dv

#for parental education
#father
m_indresp$paedu<-c(0)
m_indresp$paedu[m_indresp$paedqf%in%c(1,2,97)]<-1 #low (everything below, the category "ot
m_indresp$paedu[m_indresp$paedqf%in%c(3,4)]<-2 #middle (complete A-level or have some qual
m_indresp$paedu[m_indresp$paedqf%in%c(5)]<-3 #high (have a degree)
m_indresp$paedu[m_indresp$paedu==0]<-NA

m_indresp$paedu<-factor(m_indresp$paedu,levels=c(1,2,3),labels=c('low',"middle","high"))

#mother
m_indresp$maedu<-c(0)
m_indresp$maedu[m_indresp$maedqf%in%c(1,2,97)]<-1 #low (everything below, the category "ot
m_indresp$maedu[m_indresp$maedqf%in%c(3,4)]<-2 #middle (complete A-level or have some qual
m_indresp$maedu[m_indresp$maedqf%in%c(5)]<-3 #high (have a degree)
```

```r
m_indresp$maedu[m_indresp$maedu==0]<-NA

m_indresp$maedu<-factor(m_indresp$maedu,levels=c(1,2,3),labels=c('low',"middle","high"))

#for whether mother working at age 14
m_indresp$mawork<-c(0)
m_indresp$mawork[m_indresp$maju==2]<-1 #notworking
m_indresp$mawork[m_indresp$maju==1]<-2 #working
m_indresp$mawork[m_indresp$maju%in%c(3,4)]<-3 #Mother deceased/Mother not living with resp
m_indresp$mawork[m_indresp$mawork==0]<-NA #missing

m_indresp$mawork<-factor(m_indresp$mawork,levels=c(1,2,3),labels=c("notworking","working",
```

```r
#preparing dependent variables for multinomial regression (wave 1)
##patterns of home-leaving
###living with parents(including those who have children)=========================
m_indresp$leave[m_indresp$lvpa!=0]<-1
##route out 1: to live with a partner ==============================
m_indresp$leave[m_indresp$lvpa==0 & #not living with parents
                        m_indresp$m_mastat_dv%in%c(2,3,10)  #married
                        ]<-2

##route out 2: to live without a partner=====================
m_indresp$leave[m_indresp$lvpa==0 & #not living with parents
                        m_indresp$m_mastat_dv%in%c(0,1,4,5,6,7,8,9,10)
                        ]<-3
```

```r
#preparing dependent variables for multinomial regression (wave 13)
##patterns of home-leaving
m_indresp$ledes<-c(0)

###living with parents(including those who have children)=========================
m_indresp$ledes[m_indresp$lvpa!=0]<-1

##route out 1: to live with a partner(marriage) ==============================
m_indresp$ledes[m_indresp$lvpa==0 & #not living with parents
                        m_indresp$m_mastat_dv%in%c(2,3)  #married
                        ]<-2

##route out 2: to live with a partner(cohabitation) ==============================
```

```
m_indresp$ledes[m_indresp$lvpa==0 & #not living with parents
                        m_indresp$m_mastat_dv%in%c(10)  #in cohabitation
                        ]<-3

##route out 3: to live without a partner====================
###females
m_indresp$ledes[m_indresp$lvpa==0 & #not living with parents
                        m_indresp$m_mastat_dv%in%c(0,1,4,5,6,7,8,9)  #others
                        ]<-4


attr(m_indresp$ledes,"labels")<-c(
"NA:n=80,having missing values in marital status;",
"1:living with parents;",
"2:route out 1:leaving with partner (marriage) ;",
"3:route out 2:leaving with partner (cohabitation);",
"4:route out 3:leaving without a partner")


m_indresp$ledes<-factor(m_indresp$ledes,levels=c(1,2,3,4),labels=c("living with parents","
```

Model building with imputations

dealing with haven issues

```
a_indrespt<-a_indresp
a_indrespty<-a_indresp[which(a_indresp$a_age_dv%in%c(16:35)),]
a_indresptf<-a_indrespty[which(a_indrespty$a_sex_dv%in%c(2)),]

m_indrespt<-m_indresp
m_indrespty<-m_indresp[which(m_indresp$m_age_dv%in%c(16:35)),]
m_indresptf<-m_indrespty[which(m_indrespty$m_sex_dv%in%c(2)),]
```

```
#not ordered: mastat_dv,sex_dv,urban_dv,lvpab,racel,ledes,leave,mawork
#wave 1
a_indrespt<-a_indrespt%>%dplyr::select(a_indinus_xw,a_strata,a_psu,a_mastat_dv,a_sex_dv,a_

#as numeric
a_indrespt$a_indinus_xw<-haven::zap_labels(a_indrespt$a_indinus_xw)
```

```
a_indrespt$lnincome<-haven::zap_labels(a_indrespt$lnincome)
a_indrespt$fasesXage<-haven::zap_labels(a_indrespt$fasesXage)
a_indrespt$fases<-haven::zap_labels(a_indrespt$fases)
a_indrespt$mases<-haven::zap_labels(a_indrespt$mases)


#as factor
a_indrespt$a_mastat_dv<-haven::as_factor(a_indrespt$a_mastat_dv)
a_indrespt$a_strata<-haven::as_factor(a_indrespt$a_strata)
a_indrespt$a_psu<-haven::as_factor(a_indrespt$a_psu)
a_indrespt$a_nchild_dv<-haven::as_factor(a_indrespt$a_nchild_dv)
a_indrespt$lvpab<-haven::as_factor(a_indrespt$lvpab)
a_indrespt$racel<-haven::as_factor(a_indrespt$racel)
a_indrespt$ledes<-haven::as_factor(a_indrespt$ledes)
a_indrespt$leave<-haven::as_factor(a_indrespt$leave)
a_indrespt$mawork <-haven::as_factor(a_indrespt$mawork)
a_indrespt$edu<-haven::as_factor(a_indrespt$edu)
a_indrespt$age_group<-haven::as_factor(a_indrespt$age_group)
a_indrespt$paedu <-haven::as_factor(a_indrespt$paedu )
a_indrespt$maedu  <-haven::as_factor(a_indrespt$maedu )
a_indrespt$a_urban_dv<-haven::as_factor(a_indrespt$a_urban_dv)
a_indrespt$a_sex_dv<-haven::as_factor(a_indrespt$a_sex_dv)

#wave 13 a_indinus_xw
m_indrespt<-m_indrespt%>%dplyr::select(m_indpxui_xw,m_strata,m_psu,m_mastat_dv,m_sex_dv,m_

#as numeric
m_indrespt$m_indpxui_xw<-haven::zap_labels(m_indrespt$m_indpxui_xw)
m_indrespt$lnincome<-haven::zap_labels(m_indrespt$lnincome)
m_indrespt$fasesXage<-haven::zap_labels(m_indrespt$fasesXage)
m_indrespt$fases<-haven::zap_labels(m_indrespt$fases)
m_indrespt$mases<-haven::zap_labels(m_indrespt$mases)

#m_sex_dv,m_urban_dv,lvpab
#as factor
m_indrespt$m_mastat_dv<-haven::as_factor(m_indrespt$m_mastat_dv)
m_indrespt$m_strata<-haven::as_factor(m_indrespt$m_strata)
m_indrespt$m_psu<-haven::as_factor(m_indrespt$m_psu)
m_indrespt$m_nchild_dv<-haven::as_factor(m_indrespt$m_nchild_dv)
m_indrespt$lvpab<-haven::as_factor(m_indrespt$lvpab)
m_indrespt$racel<-haven::as_factor(m_indrespt$racel)
```

```
m_indrespt$ledes<-haven::as_factor(m_indrespt$ledes)
m_indrespt$leave<-haven::as_factor(m_indrespt$leave)
m_indrespt$mawork <-haven::as_factor(m_indrespt$mawork)
m_indrespt$edu<-haven::as_factor(m_indrespt$edu)
m_indrespt$age_group<-haven::as_factor(m_indrespt$age_group)
m_indrespt$paedu <-haven::as_factor(m_indrespt$paedu )
m_indrespt$maedu  <-haven::as_factor(m_indrespt$maedu )
m_indrespt$m_urban_dv<-haven::as_factor(m_indrespt$m_urban_dv)
m_indrespt$m_sex_dv<-haven::as_factor(m_indrespt$m_sex_dv)
```

```
#droplist: droplist, where you can pass a character vector of predictor variables that you
#https://stats.stackexchange.com/questions/209811/how-to-improve-running-time-for-r-mice-d
ImputeData <- function(data, m = 5, maxit = 5,method=NULL, droplist = NULL) {
  if (length(intersect(names(data), droplist)) < length(droplist)) {
    stop("Droplist variables not found in data set")
  }
  predictorMatrix <- (1 - diag(1, ncol(data)))
  for (term in droplist) {
  drop.index <- which(names(data) == term)
    predictorMatrix[, drop.index] <- 0
  }
  mids.out <- mice(data, m = m, maxit = maxit,method=method,
                   predictorMatrix = predictorMatrix)
  return(mids.out)
}

#Time difference of 39.56778 mins Number of logged events: 374
start.time <- Sys.time()
imp_m<-ImputeData(m_indrespt,m=1,maxit = 0,droplist = c("m_indpxui_xw","m_strata","m_psu",
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 16.62547 secs
```

```
#Time difference of 1.174014 hours Number of logged events: 290
start.time <- Sys.time()
imp_a<-ImputeData(a_indrespt,m=1,maxit = 0,droplist = c("a_indinus_xw","a_strata"  , "a_ps
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 34.7621 secs
```

```r
#adjusting the methods
impmethod_m<-imp_m$method
impmethod_m["m_sex_dv"]<-"logreg"
impmethod_m["m_urban_dv"]<-"logreg"
impmethod_m["lvpab"]<-"logreg"

impmethod_a<-imp_a$method
impmethod_a["a_sex_dv"]<-"logreg"
impmethod_a["a_urban_dv"]<-"logreg"
impmethod_a["lvpab"]<-"logreg"

#Time difference of  35.07069 mins,Number of logged events: 373
start.time <- Sys.time()
imp_m<-ImputeData(m_indrespt,m=5,maxit = 5,droplist = c("m_indpxui_xw","m_strata","m_psu",
```

```
Warning: Type mismatch for variable(s): m_sex_dv
Imputation method logreg is not for factors with >2 levels.

Warning: Type mismatch for variable(s): m_urban_dv
Imputation method logreg is not for factors with >2 levels.


 iter imp variable
  1   1  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
  1   2  m_mastat_dv  m_sex_dv  m_urban_dv  mases  edu  age_group  lnincome  fasesXage  paedu
  1   3  m_mastat_dv  m_sex_dv  m_urban_dv  mases  edu  age_group  lnincome  fasesXage  paedu
  1   4  m_mastat_dv  m_sex_dv  m_urban_dv  mases  edu  age_group  lnincome  fasesXage*  paed
  1   5  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
  2   1  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
  2   2  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
  2   3  m_mastat_dv  m_sex_dv  m_urban_dv  mases  edu  age_group  lnincome*  fasesXage  paed
  2   4  m_mastat_dv  m_sex_dv  m_urban_dv  mases  edu  age_group  lnincome  fasesXage*  paed
  2   5  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage*  pa
  3   1  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
  3   2  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome  fasesXage*  pae
  3   3  m_mastat_dv  m_sex_dv  m_urban_dv  mases  edu  age_group  lnincome  fasesXage*  paed
  3   4  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome  fasesXage  paed
  3   5  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
  4   1  m_mastat_dv  m_sex_dv  m_urban_dv  mases  edu  age_group  lnincome  fasesXage*  paed
```

```
  4   2  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage*  pa
  4   3  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome  fasesXage  paed
  4   4  m_mastat_dv  m_sex_dv  m_urban_dv  mases  edu  age_group  lnincome*  fasesXage  paed
  4   5  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage*  pa
  5   1  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage*  pa
  5   2  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage*  pa
  5   3  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
  5   4  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
  5   5  m_mastat_dv  m_sex_dv  m_urban_dv  mases*  edu  age_group  lnincome*  fasesXage  pae
```

```
Warning: Number of logged events: 368
```

```
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 36.58826 mins
```

```
#Time difference of 1.121355 hours Number of logged events: 291
start.time <- Sys.time()
imp_a<-ImputeData(a_indrespt,m=5,maxit = 5,droplist = c("a_indinus_xw","a_strata"  , "a_ps
```

```
Warning: Type mismatch for variable(s): a_sex_dv
Imputation method logreg is not for factors with >2 levels.

Warning: Type mismatch for variable(s): a_urban_dv
Imputation method logreg is not for factors with >2 levels.
```

```
 iter imp variable
  1   1  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome*  paedu  maedu  mawork  leave
  1   2  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome*  paedu  maedu  mawork  leave
  1   3  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  1   4  a_mastat_dv  a_sex_dv  mases*  edu  age_group  lnincome  paedu  maedu  mawork  leave
  1   5  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  2   1  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  2   2  a_mastat_dv  a_sex_dv  mases*  edu  age_group  lnincome*  paedu  maedu  mawork  leav
  2   3  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  2   4  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
```

```
  2   5  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  3   1  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  3   2  a_mastat_dv  a_sex_dv  mases*  edu  age_group  lnincome  paedu  maedu  mawork  leave
  3   3  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  3   4  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  3   5  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  4   1  a_mastat_dv  a_sex_dv  mases*  edu  age_group  lnincome  paedu  maedu  mawork  leave
  4   2  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome*  paedu  maedu  mawork  leave
  4   3  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  4   4  a_mastat_dv  a_sex_dv  mases*  edu  age_group  lnincome*  paedu  maedu  mawork  leav
  4   5  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  5   1  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  5   2  a_mastat_dv  a_sex_dv  mases*  edu  age_group  lnincome  paedu  maedu  mawork  leave
  5   3  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  5   4  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
  5   5  a_mastat_dv  a_sex_dv  mases  edu  age_group  lnincome  paedu  maedu  mawork  leave
```

```
Warning: Number of logged events: 286
```

```r
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 1.267352 hours
```

Wave 1 females

```r
#combine multiple imputation and design effects: function built from https://gist.github.c

lm_svy_af <- function(formula, imputations) {
  options(survey.lonely.psu="adjust")

  #setting up null objects allows us to easily add results
  #later
  b <- se<- NULL

  #now loop through our imputations and run the model
  for(i in 1:imputations$m) {
    #grab the complete dataset
    imputation <- complete(imputations, i)
```

```
    #create the design effect object
    imputation.svy <- subset(svydesign(id=~a_psu, strata=~a_strata,
                          weights=~a_indinus_xw, data=imputation),age_group%in%c("16-20","2
    #run the model
    model <- svyglm(formula, design=imputation.svy,family=quasibinomial)
    #collect the results
    b <- cbind(b, coef(model))
    se <- cbind(se, summary(model)$coef[,2])
  }

  #now pool the results
  b.pool <- apply(b, 1, mean)
  between.var <- apply(b, 1, var)
  within.var <- apply(se^2, 1, mean)
  se.pool <- sqrt(within.var+between.var+between.var/imputations$m)
  t.pool <- b.pool/se.pool
  pvalue.pool <- (1-pnorm(abs(t.pool)))*2
  coefficients <- data.frame(b.pool, se.pool, t.pool, pvalue.pool)

  #we can also grap n and p from the last model since
  #they should be the same across all iterations
  n <- nobs(model)
  p <- length(model$coefficients)-1

  #return everything in a list
  return(list(coef=coefficients,
              n=n))
}


# for model 2 and 3 ============================================================
lm_svy_af_2 <- function(formula, imputations) {
  options(survey.lonely.psu="adjust")

  #setting up null objects allows us to easily add results
  #later
  b <- se<- NULL

  #now loop through our imputations and run the model
  for(i in 1:imputations$m) {
    #grab the complete dataset
```

```
    imputation <- complete(imputations, i)
    #create the design effect object
    imputation.svy <- subset(svydesign(id=~a_psu, strata=~a_strata,
                          weights=~a_indinus_xw, data=imputation),age_group%in%c("16-20","2
    #run the model
    model <- svy_vglm(formula, design=imputation.svy,family=multinomial(refLevel=1))
    #collect the results
    b <- cbind(b, coef(model))
    se <- cbind(se, SE(model))
  }

  #now pool the results
  b.pool <- apply(b, 1, mean)
  between.var <- apply(b, 1, var)
  within.var <- apply(se^2, 1, mean)
  se.pool <- sqrt(within.var+between.var+between.var/imputations$m)
  t.pool <- b.pool/se.pool
  pvalue.pool <- (1-pnorm(abs(t.pool)))*2
  coefficients <- data.frame(b.pool, se.pool, t.pool, pvalue.pool)



  #return everything in a list
  return(list(coef=coefficients))
}
```

```
#running models

#m1 Time difference of 5.143367 mins
start.time <- Sys.time()

m1af_imp<-lm_svy_af(lvpab~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(pa
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion
```

```
Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion
```

```r
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 4.749827 mins
```

```r
#m2 Time difference of 18.08427 mins
start.time <- Sys.time()

m2af_imp<-lm_svy_af_2(ledes~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(

end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 20.30996 mins
```

```
#m3 Time difference of 10.00591 mins
start.time <- Sys.time()

m3af_imp<-lm_svy_af_2(leave~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(

end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 10.39266 mins
```

Wave 1 males

```
#with imputation: function built from https://gist.github.com/AaronGullickson/3ccb3fdd1778

lm_svy_am <- function(formula, imputations) {
  options(survey.lonely.psu="adjust")

  #setting up null objects allows us to easily add results
  #later
  b <- se<- NULL

  #now loop through our imputations and run the model
  for(i in 1:imputations$m) {
    #grab the complete dataset
    imputation <- complete(imputations, i)
    #create the design effect object
    imputation.svy <- subset(svydesign(id=~a_psu, strata=~a_strata,
                         weights=~a_indinus_xw, data=imputation),age_group%in%c("16-20","2
    #run the model
    model <- svyglm(formula, design=imputation.svy,family=quasibinomial)
    #collect the results
    b <- cbind(b, coef(model))
    se <- cbind(se, summary(model)$coef[,2])
  }

  #now pool the results
  b.pool <- apply(b, 1, mean)
  between.var <- apply(b, 1, var)
  within.var <- apply(se^2, 1, mean)
```

```
  se.pool <- sqrt(within.var+between.var+between.var/imputations$m)
  t.pool <- b.pool/se.pool
  pvalue.pool <- (1-pnorm(abs(t.pool)))*2
  coefficients <- data.frame(b.pool, se.pool, t.pool, pvalue.pool)

  #we can also grap n and p from the last model since
  #they should be the same across all iterations
  n <- nobs(model)
  p <- length(model$coefficients)-1

  #return everything in a list
  return(list(coef=coefficients,
              n=n))
}


# for model 2 and 3 ============================================================
lm_svy_am_2 <- function(formula, imputations) {
  options(survey.lonely.psu="adjust")

  #setting up null objects allows us to easily add results
  #later
  b <- se<- NULL

  #now loop through our imputations and run the model
  for(i in 1:imputations$m) {
    #grab the complete dataset
    imputation <- complete(imputations, i)
    #create the design effect object
    imputation.svy <- subset(svydesign(id=~a_psu, strata=~a_strata,
                         weights=~a_indinus_xw, data=imputation),age_group%in%c("16-20","2
    #run the model
    model <- svy_vglm(formula, design=imputation.svy,family=multinomial(refLevel=1))
    #collect the results
    b <- cbind(b, coef(model))
    se <- cbind(se, SE(model))
  }

  #now pool the results
  b.pool <- apply(b, 1, mean)
  between.var <- apply(b, 1, var)
```

```
  within.var <- apply(se^2, 1, mean)
  se.pool <- sqrt(within.var+between.var+between.var/imputations$m)
  t.pool <- b.pool/se.pool
  pvalue.pool <- (1-pnorm(abs(t.pool)))*2
  coefficients <- data.frame(b.pool, se.pool, t.pool, pvalue.pool)



  #return everything in a list
  return(list(coef=coefficients))
}
```

```
#m1 Time difference of 5.283394 mins
start.time <- Sys.time()

m1am_imp<-lm_svy_am(lvpab~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(pa
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion
```

```
Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion
```

```
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 4.648464 mins
```

```
#m2 Time difference of 17.32093 mins
start.time <- Sys.time()

m2am_imp<-lm_svy_am_2(ledes~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(

end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 15.15325 mins
```

```
#m3 Time difference of 10.53404 mins
start.time <- Sys.time()

m3am_imp<-lm_svy_am_2(leave~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(

end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 7.209314 mins
```

Wave 13 females

```r
#with imputation: function built from https://gist.github.com/AaronGullickson/3ccb3fdd1778

lm_svy_mf <- function(formula, imputations) {
  options(survey.lonely.psu="adjust")

  #setting up null objects allows us to easily add results
  #later
  b <- se<- NULL

  #now loop through our imputations and run the model
  for(i in 1:imputations$m) {
    #grab the complete dataset
    imputation <- complete(imputations, i)
    #create the design effect object
    imputation.svy <- subset(svydesign(id=~m_psu, strata=~m_strata,
                         weights=~m_indpxui_xw, data=imputation),age_group%in%c("16-20","2
    #run the model
    model <- svyglm(formula, design=imputation.svy,family=quasibinomial)
    #collect the results
    b <- cbind(b, coef(model))
    se <- cbind(se, summary(model)$coef[,2])
  }

  #now pool the results
  b.pool <- apply(b, 1, mean)
  between.var <- apply(b, 1, var)
  within.var <- apply(se^2, 1, mean)
  se.pool <- sqrt(within.var+between.var+between.var/imputations$m)
  t.pool <- b.pool/se.pool
  pvalue.pool <- (1-pnorm(abs(t.pool)))*2
  coefficients <- data.frame(b.pool, se.pool, t.pool, pvalue.pool)

  #we can also grap n and p from the last model since
  #they should be the same across all iterations
  n <- nobs(model)
  p <- length(model$coefficients)-1

  #return everything in a list
  return(list(coef=coefficients,
              n=n))
}
```

```r
# for model 2 and 3 ============================================================
lm_svy_mf_2 <- function(formula, imputations) {
  options(survey.lonely.psu="adjust")

  #setting up null objects allows us to easily add results
  #later
  b <- se<- NULL

  #now loop through our imputations and run the model
  for(i in 1:imputations$m) {
    #grab the complete dataset
    imputation <- complete(imputations, i)
    #create the design effect object
    imputation.svy <- subset(svydesign(id=~m_psu, strata=~m_strata,
                         weights=~m_indpxui_xw, data=imputation),age_group%in%c("16-20","2
    #run the model
    model <- svy_vglm(formula, design=imputation.svy,family=multinomial(refLevel=1))
    #collect the results
    b <- cbind(b, coef(model))
    se <- cbind(se, SE(model))
  }

  #now pool the results
  b.pool <- apply(b, 1, mean)
  between.var <- apply(b, 1, var)
  within.var <- apply(se^2, 1, mean)
  se.pool <- sqrt(within.var+between.var+between.var/imputations$m)
  t.pool <- b.pool/se.pool
  pvalue.pool <- (1-pnorm(abs(t.pool)))*2
  coefficients <- data.frame(b.pool, se.pool, t.pool, pvalue.pool)



  #return everything in a list
  return(list(coef=coefficients))
}
```

```r
#running models

#m1 Time difference of 2.939297 mins
```

```r
start.time <- Sys.time()

m1bf_imp<-lm_svy_mf(lvpab~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(pa
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion
```

```r
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 2.356541 mins
```

```
#m2 Time difference of 8.87294 mins
start.time <- Sys.time()

m2bf_imp<-lm_svy_mf_2(ledes~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(

end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 10.81459 mins
```

```
#m3 Time difference of 4.866161 mins
start.time <- Sys.time()

m3bf_imp<-lm_svy_mf_2(leave~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(

end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 4.761611 mins
```

Wave 13 males

```
#with imputation: function built from https://gist.github.com/AaronGullickson/3ccb3fdd1778

lm_svy_mm <- function(formula, imputations) {
  options(survey.lonely.psu="adjust")

  #setting up null objects allows us to easily add results
  #later
  b <- se<- NULL

  #now loop through our imputations and run the model
  for(i in 1:imputations$m) {
    #grab the complete dataset
    imputation <- complete(imputations, i)
    #create the design effect object
    imputation.svy <-subset(svydesign(id=~m_psu, strata=~m_strata,
```

```
                        weights=~m_indpxui_xw, data=imputation),age_group%in%c("16-20","2

    #run the model
    model <- svyglm(formula, design=imputation.svy,family=quasibinomial)
    #collect the results
    b <- cbind(b, coef(model))
    se <- cbind(se, summary(model)$coef[,2])
  }

  #now pool the results
  b.pool <- apply(b, 1, mean)
  between.var <- apply(b, 1, var)
  within.var <- apply(se^2, 1, mean)
  se.pool <- sqrt(within.var+between.var+between.var/imputations$m)
  t.pool <- b.pool/se.pool
  pvalue.pool <- (1-pnorm(abs(t.pool)))*2
  coefficients <- data.frame(b.pool, se.pool, t.pool, pvalue.pool)

  #we can also grap n and p from the last model since
  #they should be the same across all iterations
  n <- nobs(model)
  p <- length(model$coefficients)-1

  #return everything in a list
  return(list(coef=coefficients,
              n=n))
}


# for model 2 and 3 ============================================================
lm_svy_mm_2 <- function(formula, imputations) {
  options(survey.lonely.psu="adjust")

  #setting up null objects allows us to easily add results
  #later
  b <- se<- NULL

  #now loop through our imputations and run the model
  for(i in 1:imputations$m) {
    #grab the complete dataset
    imputation <- complete(imputations, i)
```

```
    #create the design effect object
    imputation.svy <-subset(svydesign(id=~m_psu, strata=~m_strata,
                          weights=~m_indpxui_xw, data=imputation),age_group%in%c("16-20","2

    #run the model
    model <- svy_vglm(formula, design=imputation.svy,family=multinomial(refLevel=1))
    #collect the results
    b <- cbind(b, coef(model))
    se <- cbind(se, SE(model))
  }

  #now pool the results
  b.pool <- apply(b, 1, mean)
  between.var <- apply(b, 1, var)
  within.var <- apply(se^2, 1, mean)
  se.pool <- sqrt(within.var+between.var+between.var/imputations$m)
  t.pool <- b.pool/se.pool
  pvalue.pool <- (1-pnorm(abs(t.pool)))*2
  coefficients <- data.frame(b.pool, se.pool, t.pool, pvalue.pool)


  #return everything in a list
  return(list(coef=coefficients))
}
```

```
#m1 Time difference of 3.179687 mins
start.time <- Sys.time()

m1bm_imp<-lm_svy_mm(lvpab~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(pa
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion
```

```r
end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 2.410123 mins
```

```r
#m2 Time difference of 8.01764 mins
start.time <- Sys.time()

m2bm_imp<-lm_svy_mm_2(ledes~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(

end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 9.423681 mins
```

```r
#m3 Time difference of 4.359161 mins
start.time <- Sys.time()
```

```
m3bm_imp<-lm_svy_mm_2(leave~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(

end.time <- Sys.time()
time.taken <- end.time - start.time
time.taken
```

```
Time difference of 5.077521 mins
```

Plotting results

```
convertModel <- function(model) {
  tr <- createTexreg(
    coef.names = rownames(model$coef),
    coef = model$coef$b.pool,
    se = model$coef$se.pool,
    pvalues = model$coef$pvalue.pool,
  )
}

#omit<-c("(Intercept):1", "(Intercept):2", "age_group21-25:1","age_group21-25:2","age_grou

#compare males and females in wave 1
 af_m1<-screenreg(lapply(list(m1af_imp, m1am_imp), convertModel))
 af_m2<-screenreg(lapply(list(m2af_imp, m2am_imp), convertModel))
 af_m3<-htmlreg(lapply(list(m3af_imp, m3am_imp), convertModel),file="compare males and fem
```

```
The table was written to the file 'compare males and females in wave 1(2).html'.
```

```
#compare males and females in wave 13
 mf_m1<-screenreg(lapply(list(m1bf_imp, m1bm_imp), convertModel))
 mf_m2<-screenreg(lapply(list(m2bf_imp, m2bm_imp), convertModel))
 mf_m3<-htmlreg(lapply(list(m3bf_imp, m3bm_imp), convertModel),file="compare males and fem
```

```
The table was written to the file 'compare males and females in wave 13(2).html'.
```

```
#compare females in wave 13 and wave 1
 bf_m1<-screenreg(lapply(list(m1af_imp, m1bf_imp), convertModel))
```

```
bf_m2<-screenreg(lapply(list(m2af_imp, m2bf_imp), convertModel))
bf_m3<-htmlreg(lapply(list(m3af_imp, m3bf_imp), convertModel),file="compare females in wa
```

```
The table was written to the file 'compare females in wave 13 and wave 1(2).html'.
```

```
#compare males in wave 13 and wave 1
bm_m1<-screenreg(lapply(list(m1am_imp, m1bm_imp), convertModel))
bm_m2<-screenreg(lapply(list(m2am_imp, m2bm_imp), convertModel))
bm_m3<-htmlreg(lapply(list(m3am_imp, m3bm_imp), convertModel),file="compare males in wave
```

```
The table was written to the file 'compare males in wave 13 and wave 1(2).html'.
```

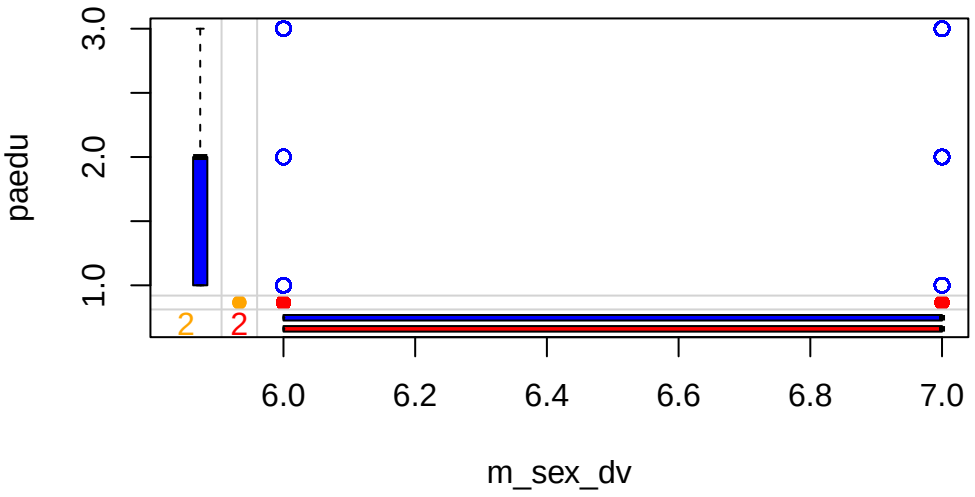
Assumption checking

imputation

https://stats.oarc.ucla.edu/r/faq/how-do-i-perform-multiple-imputation-using-predictive-mean-matching-in-r/

```
a_indrespy<-a_indresp[which(a_indresp$a_age_dv%in%c(16:35)),]
m_indrespy<-m_indresp[which(m_indresp$m_age_dv%in%c(16:35)),]
```

```
marginplot(m_indrespt[c(5, 16)], col = c("blue", "red", "orange"))
```

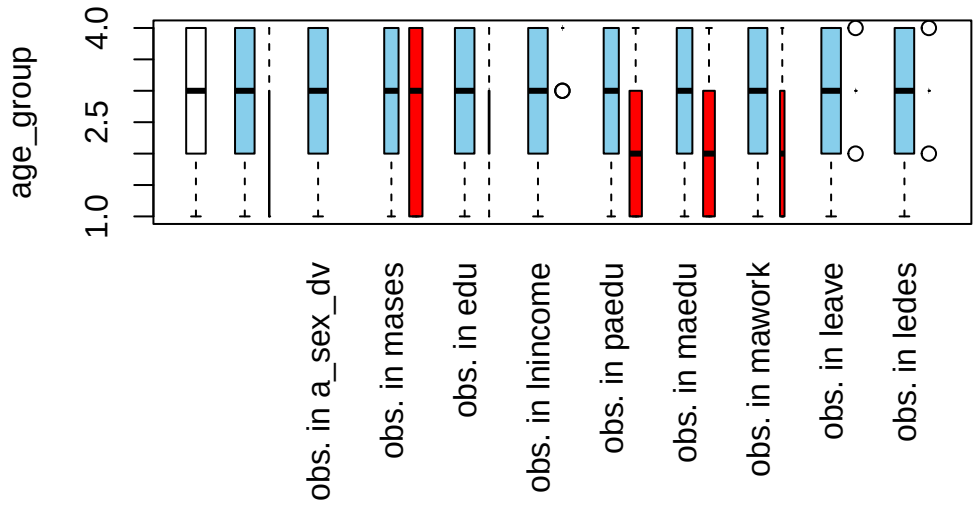


The box plots are blue for observed values and red for missing values

Evidence supporting MAR over MCAR would be if the distribution of **x2** for those observations with missing information for **y1** or **y4** were much higher or much lower than those of the non-missing observations. In the above graph, the boxplots appear to mostly overlap once again providing support for the assumption of MCAR.

```
pbox(a_indrespt, pos = 11)
```

```
Warning in createPlot(main, sub, xlab, ylab, labels, ca$at): not enough space
to display frequencies
```



Models

Model building without imputations

Adopt the Weightings

```
options(survey.lonely.psu="adjust")

svy_indresp1 <- svydesign(id=~a_psu, strata=~a_strata,
                          weights=~a_indinus_xw, data=a_indresp)

svy_indresp2 <- svydesign(id=~m_psu, strata=~m_strata,weights=~m_indpxui_xw, data=m_indres

#subset the data for respondents aged 25-50
svy_indresp1y<-subset(svy_indresp1, a_age_dv%in%c(16:35))

svy_indresp2y<-subset(svy_indresp2, m_age_dv%in%c(16:35))


#for males and females
```

```
#females
svy_indresp1yf<-subset(svy_indresp1y, a_sex_dv%in%c(2))
svy_indresp2yf<-subset(svy_indresp2y, m_sex_dv%in%c(2))


#males
svy_indresp1ym<-subset(svy_indresp1y, a_sex_dv%in%c(1))
svy_indresp2ym<-subset(svy_indresp2y, m_sex_dv%in%c(1))
```

Model 1

```
#for females
#without imputation
#females
m1af<-svyglm(lvpab~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+fa
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(glm.object): observations with zero weight not used for
calculating dispersion
```

```
m1bf<-svyglm(lvpab~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+fa
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion
```

```
#males
m1am<-svyglm(lvpab~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+fa
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion

Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion
```

```
m1bm<-svyglm(lvpab~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+fa
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion
```

```
Warning in summary.glm(g): observations with zero weight not used for
calculating dispersion
```

```
#z tests of coefficients
zscore <- function(coef2,coef1,se2,se1){(coef2-coef1)/sqrt((se2)**2+(se1)**2)
}
#females
#fases
zscore(coef2 =    1.35783  , coef1=  0.493502       , se2= 0.46276 , se1= 0.084185 )
```

```
[1] 1.837607
```

```
#maedu (high)
zscore(coef2 = -0.16265      ,coef1= -0.484515      , se2=0.53702   , se1=0.202592)
```

```
[1] 0.5607762
```

```
#paedu (high)
zscore(coef2 =   0.50519       ,coef1=  -0.200648       , se2=0.57399   , se1=0.185489  )
```

```
[1] 1.170123
```

```
#mawork sig
zscore(coef2 =  -0.38087             , coef1=  0.359646         , se2= 0.31075    , se1= 0
```

```
[1] -2.239524
```

```
#males
#fases
zscore(coef2 =  0.77227           ,coef1=   0.292779         , se2=  0.66703 , se1=  0.093
```

```
[1] 0.7118632
```

```
#maedu (high)
zscore(coef2 =    -0.79371           ,coef1=-0.188297        , se2=0.63461    , se1=0.21855
```

```
[1] -0.9020002
```

```
#paedu (high)
zscore(coef2 =   -0.19252            ,coef1= -0.320714          , se2=0.58333    , se1=0.2
```

```
[1] 0.2058279
```

```
#mawork sig
zscore(coef2 =  -0.63024             , coef1= 0.181253           , se2= 0.35183     , se1=
```

```
[1] -2.16006
```

```
#ames
#females difference -0.122  p.value  0.021
cames<-summary(marginaleffects(m1af,variables="mawork"))
```

```
Warning: This function has been renamed to `slopes()`. The `marginaleffects()`
  alias will be removed in the near future.
```

```
cames2<-summary(marginaleffects(m1bf,variables="mawork"))
```

```
Warning: This function has been renamed to `slopes()`. The `marginaleffects()`
  alias will be removed in the near future.
```

```
came<-rbind(cames,cames2)
compare.margins(came$estimate,came$std.error)
```

```
$difference
[1] -0.122

$p.value
[1] 0.021
```

```
#males difference -0.141; p.value  0.017
cames<-summary(marginaleffects(m1am,variables="mawork"))
```

```
Warning: This function has been renamed to `slopes()`. The `marginaleffects()`
  alias will be removed in the near future.
```

```
cames2<-summary(marginaleffects(m1bm,variables="mawork"))
```

```
Warning: This function has been renamed to `slopes()`. The `marginaleffects()`
  alias will be removed in the near future.
```

```
came<-rbind(cames,cames2)
compare.margins(came$estimate,came$std.error)
```

```
$difference
[1] -0.141

$p.value
[1] 0.017
```

Model 2

```
m2af<-svy_vglm(ledes~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+

m2bf<-svy_vglm(ledes~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 2
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 3
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 5
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 6
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 6
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 6
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12


Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 7
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12


Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 8
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 8
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 8
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 8
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
m2am<-svy_vglm(ledes~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+

m2bm<-svy_vglm(ledes~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 1
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 1
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 2
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 6
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 6
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
#z test for coefficients
#females
##cohabitation
###pases
zscore(coef2 =  -1.145945              ,coef1=   -0.5279231          , se2=  0.578323 , s
```

```
[1] -1.050876
```

```
###maedu(high)
zscore(coef2 =  0.535709             ,coef1=    0.3910630          , se2=  0.647512   , se
```

```
[1] 0.2096059
```

```
###paedu(high)
zscore(coef2 = -0.408988           ,coef1=   -0.0225736          , se2=  0.690724, se1= 0.2
```

```
[1] -0.5329859
```

```
###mawork
zscore(coef2 =  0.247904          ,coef1=   -0.2549268         , se2=  0.359905 , se1=  0.1
```

```
[1] 1.302469
```

```
##marriage
###pases
zscore(coef2 = -1.358326          ,coef1=   -0.7566254          , se2=  0.563529 , se1= 0.
```

```
[1] -1.040798
```

```
###maedu(high)
zscore(coef2 =   0.358469           ,coef1=    0.2141034         , se2=  0.612275 , se1=
```

```
[1] 0.2191313
```

```
###paedu(high)
zscore(coef2 =   -0.120479          ,coef1=    0.3031643         , se2=  0.632659  , se1=
```

```
[1] -0.6321739
```

```
###mawork sig=1.97 not sig coefficient
zscore(coef2 = 0.391774           ,coef1=  -0.3863171          , se2=  0.371241 , se1=  0.
```

```
[1] 1.970451
```

```
##without a partner
###pases
zscore(coef2 =  -1.604960        ,coef1=    -0.4148343          , se2=  0.626037 , se1= 0
```

```
[1] -1.878401
```

```
###maedu(high)
zscore(coef2 =    -0.165066          ,coef1=    0.6751570        , se2=   0.606883 , se1=
```

```
[1] -1.302938
```

```
###paedu(high)
zscore(coef2 =    -0.915336          ,coef1=     0.2940644          , se2=   0.634411  , s
```

```
[1] -1.814797
```

```
###mawork sig=1.97 sig cof for wave 1, not sig for wave 13
zscore(coef2 =  0.384831           ,coef1=  -0.4305743          , se2=  0.394867 , se1=  0
```

```
[1] 1.970733
```

```
#males
##cohabitation
###pases
zscore(coef2 = -0.1573351         ,coef1=    -0.4207956          , se2=  0.8356579  , se1
```

```
[1] 0.3105553
```

```
###maedu(high)
zscore(coef2 =    0.7581093             ,coef1=    0.0648927       , se2=   0.8468744 , s
```

```
[1] 0.7833057
```

```
###paedu(high)
zscore(coef2 =    -0.0895241          ,coef1=      -0.1840590         , se2=    0.7805476
```

```
[1] 0.1147835
```

```
###mawork sig=2.09847 not sig coefficient
zscore(coef2 =   1.0827787             ,coef1=  -0.1310800          , se2=   0.5570283  ,
```

```
[1] 2.09847
```

```
##marriage
###pases
zscore(coef2 = -0.5434834        ,coef1=     -1.0634425          , se2=  0.8331357  , se1
```

```
[1] 0.6124545
```

```
###maedu(high)
zscore(coef2 =     1.4201381               ,coef1=   0.2325718        , se2=    0.7789965
```

```
[1] 1.44462
```

```
###paedu(high)
zscore(coef2 =    -0.1181511            ,coef1=      0.1469352         , se2=    0.7352541
```

```
[1] -0.3421938
```

```
###mawork
zscore(coef2 =    0.3984417            ,coef1=  -0.0598869            , se2=  0.4306555    ,
```

```
[1] 1.001882
```

```
##without a partner
###pases
zscore(coef2 = -1.1276304         ,coef1=     -0.1783426          , se2=  0.7328520 , se1=
```

```
[1] -1.282927
```

```
###maedu(high)
zscore(coef2 =      0.4877030               ,coef1=    0.2378884         , se2=     0.726
```

```
[1] 0.3260185
```

```
###paedu(high)
zscore(coef2 =     0.6922228            ,coef1=       0.6772597          , se2=    0.684850
```

```
[1] 0.02060392
```

```
###mawork
zscore(coef2 =    0.5349542             ,coef1= -0.3293337            , se2=  0.4310773
```

```
[1] 1.89286
```

```
#m2af sig
#fases:1,2,3 -
#factor(mawork)working:1,3 -
#factor(maedu)high:3 +


#m2bf sig
#fases:1,2,3 -

#m2am sig
#fases:1,2,3 -
#factor(mawork)working:3 -

#m2bm sig none
```

Model 3

```
m3af<-svy_vglm(leave~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+

m3bf<-svy_vglm(leave~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 2
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 2
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12


Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 4
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 4
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 4
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 4
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 4
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12


Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 5
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 5
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 5
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 5
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
m3am<-svy_vglm(leave~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+

m3bm<-svy_vglm(leave~fases+fasesXage+age_group+factor(mawork)+factor(maedu)+factor(paedu)+
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 1
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12

Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 1
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12


Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 2
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
Warning in checkwz(wz, M = M, trace = trace, wzepsilon = control$wzepsilon): 3
diagonal elements of the working weights variable 'wz' have been replaced by
1.819e-12
```

```
#z test for coefficients
#females
##with partners
###pases
zscore(coef2 =     -1.4293392                  ,coef1=  -0.4570019             , se2=  0.5
```

```
[1] -1.820463
```

```
###maedu(high)

###paedu(high)

###mawork
```

```
#m3bm sig
#mawork 2 +

#m3am sig
#fases 1,2 -


#m3bf
#fases 1,2 -

#m3af
#fases 1,2-
#mawork 1,2-
#maedu high +
```

```
#whether mother working when male respondents age 14 have higher predicability of home-lea
#mawork sig at wave 13 not wave 1
zscore(coef2=   0.799274      ,coef1= -0.2217087        ,se2= 0.395233   ,se1= 0.1357911)
```

```
[1] 2.443071
```

```r
#for females the effect direction is reversed, with effects only sig at wave 1
zscore(coef2=    0.3330332      ,coef1=  -0.3454488         ,se2= 0.3347052   ,se1=  0.117
```

```
[1] 1.913116
```

```r
testa<-a_indresptf%>%dplyr::select(leave,fases,fasesXage,age_group,mawork,maedu,paedu,edu,
test1<-complete.cases(testa)
table(test1)
```

```
test1
FALSE  TRUE
 3615  5690
```

```r
testb<-m_indrespt%>%dplyr::select(leave,fases,fasesXage,age_group,mawork,maedu,paedu,edu,r
test2<-complete.cases(testb)
table(test2)
```

```
test2
FALSE  TRUE
26833  1165
```

```r
#comple cases of males and females in wave 1 and 13: 434, 3927
```