

## ASDA3 Analysis Examples Replication Chapter 11

### R Code

```
# Use data sets previously prepared in SAS for this chapter to reduce code burden in R

## One Wave Analyses
# Complete Case 1 Wave
# 11.3.1 Example: Descriptive Estimation at a Single Wave, Complete Case Analysis Table 11.2

library(survey)
library(haven)
library(mice)

hrs_1wave <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/cc_1wave.sas7bdat")
names(hrs_1wave)

svyhrs_cc_1 <- svydesign(strata=~stratum, id=~secu, weights=~kgwtr, data=hrs_1wave, nest=T)
ex11_1 <- svymean(~ln_inc08, design=svyhrs_cc_1, se=T, ci=T, keep.vars=T, na.rm=T)
# Exponent of Mean, se, and CI'S
exp(ex11_1)
exp(confint(ex11_1))

# Adjusted Weight 1 Wave
hrs_1wave_adj <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/adj_wgt_1wave.sas7bdat")
names(hrs_1wave_adj)

# Set design
svyhrs_adj_1 <- svydesign(strata=~stratum, id=~secu, weights=~adj_kwgtr, data=hrs_1wave_adj, nest=T)
ex11_1_adj <- svymean(~ln_inc08, design=svyhrs_adj_1, se=T, ci=T, keep.vars=T, na.rm=T)

# Exponent of Mean, se, and CI'S
exp(ex11_1_adj)
exp(confint(ex11_1_adj))

# Multiple Imputation 1 Wave
# Use SAS data set already prepared for this example
# Load needed packages for MI, mice already loaded but also need mitools

library(mitools)

b <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/wt_deciles_1.sas7bdat")
names(b)
summary(b)

b$selfrhealth_06 <- factor(b$selfrhealth_06)
b$marcat_06 <- factor(b$marcat_06)
b$racecat <- factor(b$racecat)
b$edcat <- factor(b$edcat)
b$stratum <- factor(b$stratum)
b$kgwtr_dec <- as.numeric(b$kgwtr_dec)
summary(b)

# subset variables by number position in data set
(hrs_mi_1wave_sub <- b[, c(10,11,13,41,42,44,45,46,65,66,67,68,71)])
summary(hrs_mi_1wave_sub)

# use mice to examine and impute missing data
md.pattern(hrs_mi_1wave_sub)
```

```

ini <- mice(hrs_mi_1wave_sub, maxiter=0)
summary(ini)

# add a predictor matrix to control imputation model predictors for each imputed variable
pred <- ini$predictorMatrix
pred[,"kwgtr"] <- 0
pred[,"secu"] <- 0
pred[,"kwgtr_dec"] <- 1
pred

imphrs1wave <- mice(hrs_mi_1wave_sub, m=5, pred=pred, seed=41279)
(imphrs1wave)
summary(imphrs1wave)

# combine all five data.frame objects into a list
list_of_data_frames <-
  list(
    mice::complete( imphrs1wave , 1 ) ,
    mice::complete( imphrs1wave , 2 ) ,
    mice::complete( imphrs1wave , 3 ) ,
    mice::complete( imphrs1wave , 4 ) ,
    mice::complete( imphrs1wave , 5 )
  )

# coerce weight variable to numeric (coercing directly factor to numeric here changes the levels)
list_of_data_frames <-
  lapply(list_of_data_frames ,
    function( w ) transform( w , kwgtr_dec = as.numeric( as.character( kwgtr_dec ) ) )
  )

# multiply-imputed design statement
deshrs_1wave <-
  svydesign(
    id = ~ secu ,
    strat = ~ stratum ,
    weight = ~ kwgtr_dec ,
    data = imputationList(list_of_data_frames ) ,
    nest = TRUE
  )

# Multiply-imputed average & SE of that variable
MIcombine(with(deshrs_1wave , svymean( ~ ln_inc08 ) ) )

summary(m1w)
exp(10.41422) #mean
exp(10.36371) #lci
exp(10.46473) #uci

# multiple Imputation using a Selection Model Not Available in R

## 2 Waves of Data

# Complete Case 2 Waves
hrs_2wave <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/cc_2waves.sas7bdat")

```

```

names(hrs_2wave)

svyhrs_cc_2 <- svydesign(strata=~stratum, id=~secu, weights=~kwgtr, data=hrs_2wave, nest=T)

ex11_2 <- svymean(~incdiff_06_10, design=svyhrs_cc_2, se=T, ci=T, keep.vars=T, na.rm=T)
show(ex11_2)
confint(ex11_2)

# Adjusted Weight 2 Wave
hrs_2waves_adj <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/adj_wgt_2waves.sas7bdat")
names(hrs_2waves_adj)

svyhrs_adj_1 <- svydesign(strata=~stratum, id=~secu, weights=~adj_kwgtr, data=hrs_2waves_adj, nest=T)

show(ex11_2_adj <- svymean(~incdiff_06_10, design=svyhrs_adj_1, se=T, ci=T, keep.vars=T, na.rm=T))
confint(ex11_2_adj)

# Multiple Imputation for 2 Waves, SAS data set already prepared for this example
hrs_a <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/wt_deciles_2.sas7bdat")
names(hrs_a)

hrs_a$selfrhealth_06 <- factor(hrs_a$selfrhealth_06)
hrs_a$marcat_06 <- factor(hrs_a$marcat_06)
hrs_a$racecat <- factor(hrs_a$racecat)
hrs_a$edcat <- factor(hrs_a$edcat)
hrs_a$STRATUM <- factor(hrs_a$STRATUM)
hrs_a$kwgtr_dec <- factor(hrs_a$kwgtr_dec)

# Subset variables by number position in data set, keep those needed for imputation
hrs_mi_2waves_sub <- hrs_a [, c(10,11,13,41,42,44,45,46,65,66,67,69,71)]
names(hrs_mi_2waves_sub)

# Use mice to impute missing data
library(mice)

# Dry run to prepare the predictor matrix
ini <- mice(hrs_mi_2waves_sub, maxiter=0)
summary(ini)

# Add a predictor matrix to control imputation model predictors for each imputed variable
pred <- ini$predictorMatrix
pred[,"KWGTR"] <- 0
pred[,"SECU"] <- 0
pred[,"kwgtr_dec"] <- 1
pred[,"ln_inc06"] <- 1
pred

# Use same variables as from C11 Stata example
imphrs2waves1 <- mice(hrs_mi_2waves_sub, pred=pred, m=5, seed=41279, method="norm.nob", print=FALSE)
imphrs2waves1

# Combine all five data.frame objects into a list
list_of_data_frames <-
  list(
    mice::complete( imphrs2waves1 , 1 ) ,
    mice::complete( imphrs2waves1 , 2 ) ,
    mice::complete( imphrs2waves1 , 3 ) ,

```

```

        mice::complete( imphrs2waves1 , 4 ) ,
        mice::complete( imphrs2waves1 , 5 )
    )

# Coerce weight variable to numeric
list_of_data_frames <-
  lapply(
    list_of_data_frames ,
    function( w ) transform( w , kwgtr_dec = as.numeric( as.character( kwgtr_dec ) ) )
  )

# Set survey design , use 2006 individual weight for survey design setup
# Multiply-imputed design statement
deshrs_2waves <-
  svydesign(
    id = ~ SECU ,
    strat = ~ STRATUM ,
    weight = ~ KWGTR ,
    data = imputationList( list_of_data_frames ) ,
    nest = TRUE
  )
deshrs_2waves

# Control values used in analysis
deshrs_2waves <- update(deshrs_2waves, ln_inc10a=ifelse(ln_inc10 > 14.92, 14.92, ln_inc10),
inc10=exp(ln_inc10a), inc06=exp(ln_inc06))
deshrs_2waves <- update(deshrs_2waves, new_chg0610=(inc10 - inc06))
deshrs_2waves <- update(deshrs_2waves, new_chg0610a=ifelse(new_chg0610 < -12300000, -12300000, new_chg0610),
new_chg0610b=ifelse(new_chg0610a > 2062968, 2062968, new_chg0610a))
hrs_2w_meandiff <- with(deshrs_2waves, svymean(~(new_chg0610b), se=T, ci=T ))
hrs_2w_meandiff

# Use MIcombine for overall combined and design-adjusted mean/se
(hrs_2w_comb <- MIcombine(hrs_2w_meandiff))
(hrs_2w_comb)
confint(hrs_2w_comb)

# No Selection Model Multiple Imputation Available in R

# Calibration Method
hrs_2waves_cal <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R
Code/calibration_2waves.sas7bdat")
names(hrs_2waves_cal)

# Subset of cases without missing weight variable
hrs_2waves_calsub <- hrs_2waves_cal[ which(hrs_2waves_cal$kwgtr_cal > 0),]
summary(hrs_2waves_calsub$kwgtr_cal)

# Get mean diff 06-10
svyhrs_cal_2 <- svydesign(strata=~stratum, id=~secu, weights=~kwgtr_cal, data=hrs_2waves_calsub, nest=T,
na.rm=T)
show(ex11_2_cal <- svymean(~incdiff_06_10, design=svyhrs_cal_2, se=T, ci=T, keep.vars=T, na.rm=T))
confint(ex11_2_cal)

```

```

## 3+ Waves of Data
# Example 11.3.3 Weighted Multilevel Modeling not available in R Survey Package

# Example 11.3.3.1 Veiga Method for multi-level modeling not available in R Survey Package

# Example 11.3.4 Weighted GEE Analysis using Geepack from R ( See geepack.pdf for details)
# Data set prepared using SAS data management and estimation tools, not repeated here, see either Stata or SAS
code for details

# Load geepack
library(geepack)
hrs_3w_gee <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/wgt_gee_3pwaves.sas7bdat")
names(hrs_3w_gee)

# set factor variables
hrs_3w_gee$gender <- as.factor(hrs_3w_gee$gender)
hrs_3w_gee$stratum <- as.factor(hrs_3w_gee$stratum)
hrs_3w_gee$year <- as.factor(hrs_3w_gee$year)

# model for Example 11.3.4
# model formula
mf <- formula(ln_inc~yrssince06 + gender + yrs06sq + (yrssince06*gender) + (yrs06sq*gender) + stratum)
mf

# Run model using geeglm with weight
ex11_3_4 <- geeglm(mf, data=hrs_3w_gee, id=newid_num, weight=casewt, family=gaussian("identity"),
corstr="exchangeable")
summary(ex11_3_4)

```

## R Results

```
> # ASDA3 Chapter 11 Longitudinal Analysis HRS data
> # Use data sets previously prepared in SAS for this chapter to reduce code burden in R
>
> ## One Wave Analyses
> # Complete Case 1 Wave
> # 11.3.1 Example: Descriptive Estimation at a Single Wave, Complete Case Analysis Table 11.2
>
> library(survey)
> library(haven)
> library(mice)
```

Attaching package: 'mice'

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

filter

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

cbind, rbind

Warning message:

package 'mice' was built under R version 4.4.3

```
>
> hrs_1wave <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/cc_1wave.sas7bdat")
> names(hrs_1wave)
 [1] "hhid"          "pn"            "kfinr"         "KC001"         "KC010"
 [6] "KC070"        "gender"        "hispanic"      "schlyrs"       "secu"
[11] "stratum"      "kmarst"        "kwgtr"         "KWHYORWT"     "lfinr"
[16] "lmarst"       "lwgtr"         "mfinr"         "mmarst"        "mwgtr"
[21] "nfinr"        "nmarst"        "nwgtr"         "H8ATOTA"       "H9ATOTA"
[26] "H10ATOTA"     "H11ATOTA"      "H8ITOT"        "H9ITOT"        "H10ITOT"
[31] "H11ITOT"     "LC001"         "LC010"         "LC070"         "MC001"
[36] "MC010"        "MC070"         "NC001"         "NC010"         "NC070"
[41] "marcat_06"    "diabetes_06"   "numfalls24_06" "arthritis_06"  "selfrhealth_06"
[46] "age_06"       "marcat_08"     "diabetes_08"   "numfalls24_08" "arthritis_08"
[51] "selfrhealth_08" "age_08"        "marcat_10"     "diabetes_10"   "numfalls24_10"
[56] "arthritis_10" "selfrhealth_10" "age_10"        "marcat_12"     "diabetes_12"
[61] "numfalls24_12" "arthritis_12"  "selfrhealth_12" "age_12"        "edcat"
[66] "racecat"      "ln_inc06"      "ln_inc08"      "ln_inc10"      "ln_inc12"
>
> svyhrs_cc_1 <- svydesign(strata=~stratum, id=~secu, weights=~kwgtr, data=hrs_1wave, nest=T)
> ex11_1 <- svymean(~ln_inc08, design=svyhrs_cc_1, se=T, ci=T, keep.vars=T, na.rm=T)
> # Exponent of Mean, se, and CI'S
> exp(ex11_1)
      mean      SE
ln_inc08 34224 0.0263
> exp(confint(ex11_1))
      2.5 %    97.5 %
ln_inc08 32504.8 36034.07
>
```

```

> # Adjusted Weight 1 Wave
> hrs_1wave_adj <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/adj_wgt_1wave.sas7bdat")
> names(hrs_1wave_adj)
 [1] "hhid"          "pn"            "kfinr"         "KC001"         "KC010"
 [6] "KC070"         "gender"        "hispanic"      "schlyrs"       "secu"
[11] "stratum"       "kmarst"        "kwgtr"         "KWHYORWT"     "lfinr"
[16] "lmarst"        "lwgtr"         "mfinr"         "mmarst"        "mwgtr"
[21] "nfinr"         "nmarst"        "nwgtr"         "H8ATOTA"       "H9ATOTA"
[26] "H10ATOTA"      "H11ATOTA"      "H8ITOT"        "H9ITOT"        "H10ITOT"
[31] "H11ITOT"       "LC001"         "LC010"         "LC070"         "MC001"
[36] "MC010"         "MC070"         "NC001"         "NC010"         "NC070"
[41] "marcat_06"     "diabetes_06"   "numfalls24_06" "arthritis_06"  "selfrhealth_06"
[46] "age_06"        "marcat_08"     "diabetes_08"   "numfalls24_08" "arthritis_08"
[51] "selfrhealth_08" "age_08"        "marcat_10"     "diabetes_10"   "numfalls24_10"
[56] "arthritis_10"  "selfrhealth_10" "age_10"        "marcat_12"     "diabetes_12"
[61] "numfalls24_12" "arthritis_12"  "selfrhealth_12" "age_12"        "edcat"
[66] "racecat"       "ln_inc06"      "ln_inc08"      "ln_inc10"      "ln_inc12"
[71] "resp08"        "_LEVEL_"       "phat"          "dec"           "mean_phat"
[76] "adj_kwgtr"
>
> # Set design
> svyhrs_adj_1 <- svydesign(strata=~stratum, id=~secu, weights=~adj_kwgtr, data=hrs_1wave_adj, nest=T)
> ex11_1_adj <- svymean(~ln_inc08, design=svyhrs_adj_1, se=T, ci=T, keep.vars=T, na.rm=T)
>
> # Exponent of Mean, se, and CI'S
> exp(ex11_1_adj)
      mean      SE
ln_inc08 33309 0.0266
> exp(confint(ex11_1_adj))
      2.5 %    97.5 %
ln_inc08 31616.23 35092.42
>
>
>

```

```

> # Multiple Imputation 1 Wave
> # Use SAS data set already prepared for this example
> # Load needed packages for MI, mice already loaded but also need mitools
>
> library (mitools)
>
> b <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/wt_deciles_1.sas7bdat")
> names(b)

```

- [1] "hhid" "pn" "kfinr" "KC001" "KC010"
- [6] "KC070" "gender" "hispanic" "schlyrs" "secu"
- [11] "stratum" "kmarst" "kgwtr" "KWHYORWT" "lfinr"
- [16] "lmarst" "lwgtr" "mfinr" "mmarst" "mwgtr"
- [21] "nfinr" "nmarst" "nwgtr" "H8ATOTA" "H9ATOTA"
- [26] "H10ATOTA" "H11ATOTA" "H8ITOT" "H9ITOT" "H10ITOT"
- [31] "H11ITOT" "LC001" "LC010" "LC070" "MC001"
- [36] "MC010" "MC070" "NC001" "NC010" "NC070"
- [41] "marcat\_06" "diabetes\_06" "numfalls24\_06" "arthritis\_06" "selfrhealth\_06"
- [46] "age\_06" "marcat\_08" "diabetes\_08" "numfalls24\_08" "arthritis\_08"
- [51] "selfrhealth\_08" "age\_08" "marcat\_10" "diabetes\_10" "numfalls24\_10"
- [56] "arthritis\_10" "selfrhealth\_10" "age\_10" "marcat\_12" "diabetes\_12"
- [61] "numfalls24\_12" "arthritis\_12" "selfrhealth\_12" "age\_12" "edcat"
- [66] "racecat" "ln\_inc06" "ln\_inc08" "ln\_inc10" "ln\_inc12"
- [71] "kgwtr\_dec"

```

> summary(b)

```

hhid	pn	kfinr	KC001	KC010	KC070
Length:11789	Length:11789	Min. :1	Min. :1.000	Min. :1.000	Min. :1.000
Class :character	Class :character	1st Qu.:1	1st Qu.:2.000	1st Qu.:5.000	1st Qu.:1.000
Mode :character	Mode :character	Median :1	Median :3.000	Median :5.000	Median :1.000
		Mean :1	Mean :2.918	Mean :4.145	Mean :2.492
		3rd Qu.:1	3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:5.000
		Max. :1	Max. :5.000	Max. :9.000	Max. :9.000
			NA's :18	NA's :1	NA's :1
gender	hispanic	schlyrs	secu	stratum	kmarst
Min. :1.000	Min. :1.000	Min. : 0.00	Min. :1.000	Min. : 1.00	Min. :1.000
1st Qu.:1.000	1st Qu.:5.000	1st Qu.:12.00	1st Qu.:1.000	1st Qu.:19.00	1st Qu.:1.000
Median :2.000	Median :5.000	Median :12.00	Median :2.000	Median :33.00	Median :2.000
Mean :1.564	Mean :4.678	Mean :12.52	Mean :1.503	Mean :30.63	Mean :1.924
3rd Qu.:2.000	3rd Qu.:5.000	3rd Qu.:15.00	3rd Qu.:2.000	3rd Qu.:44.00	3rd Qu.:3.000
Max. :2.000	Max. :5.000	Max. :99.00	Max. :2.000	Max. :56.00	Max. :5.000
kgwtr	KWHYORWT	lfinr	lmarst	lwgtr	mfinr
Min. : 924	Min. :0	Min. :1.000	Min. :1.000	Min. : 0	Min. :1.000
1st Qu.: 2433	1st Qu.:0	1st Qu.:1.000	1st Qu.:1.000	1st Qu.: 1930	1st Qu.:1.000
Median : 3723	Median :0	Median :1.000	Median :2.000	Median : 3447	Median :1.000
Mean : 4458	Mean :0	Mean :1.042	Mean :1.928	Mean : 4031	Mean :1.079
3rd Qu.: 5296	3rd Qu.:0	3rd Qu.:1.000	3rd Qu.:3.000	3rd Qu.: 5266	3rd Qu.:1.000
Max. :17035	Max. :0	Max. :5.000	Max. :5.000	Max. :16558	Max. :5.000
		NA's :1215	NA's :1170		NA's :2387
mmarst	mwgtr	nfinr	nmarst	nwgtr	H8ATOTA
Min. :1.000	Min. : 0	Min. :1.0	Min. :1.000	Min. : 0	Min. :-2199392
1st Qu.:1.000	1st Qu.: 1102	1st Qu.:1.0	1st Qu.:1.000	1st Qu.: 0	1st Qu.: 33000
Median :2.000	Median : 3186	Median :1.0	Median :2.000	Median : 2643	Median : 163038
Mean :1.938	Mean : 3511	Mean :1.1	Mean :1.958	Mean : 3271	Mean : 441484
3rd Qu.:3.000	3rd Qu.: 4703	3rd Qu.:1.0	3rd Qu.:3.000	3rd Qu.: 4590	3rd Qu.: 460000
Max. :5.000	Max. :19382	Max. :5.0	Max. :5.000	Max. :19853	Max. :43512000
NA's :2357		NA's :3229	NA's :3228		
H9ATOTA	H10ATOTA	H11ATOTA	H8ITOT	H9ITOT	
Min. : -850000	Min. :-1110692	Min. :-1028000	Min. : 0	Min. : 0	
1st Qu.: 32000	1st Qu.: 28000	1st Qu.: 22775	1st Qu.: 15468	1st Qu.: 16084	

Median : 164200	Median : 148500	Median : 140085	Median : 31188	Median : 32464	
Mean : 451064	Mean : 413473	Mean : 411182	Mean : 56915	Mean : 61954	
3rd Qu.: 470150	3rd Qu.: 431000	3rd Qu.: 420000	3rd Qu.: 60600	3rd Qu.: 63690	
Max. :40660592	Max. :49900000	Max. :43921209	Max. :25360250	Max. :60014376	
NA's :1215	NA's :2387	NA's :3229		NA's :1215	
H10ITOT	H11ITOT	LC001	LC010	LC070	MC001
Min. : 0	Min. : 0	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.: 16314	1st Qu.: 15163	1st Qu.:2.000	1st Qu.:5.000	1st Qu.:1.000	1st Qu.:2.000
Median : 31564	Median : 28054	Median :3.000	Median :5.000	Median :1.000	Median :3.000
Mean : 52914	Mean : 50545	Mean :2.971	Mean :4.065	Mean :2.406	Mean :2.926
3rd Qu.: 60734	3rd Qu.: 53692	3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:4.000
Max. :3025280	Max. :3666240	Max. :5.000	Max. :9.000	Max. :9.000	Max. :5.000
NA's :2387	NA's :3229	NA's :1225	NA's :1215	NA's :1215	NA's :2395
MC010	MC070	NC001	NC010	NC070	marcat_06
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000
Median :5.000	Median :1.000	Median :3.000	Median :5.000	Median :1.000	Median :2.000
Mean :3.963	Mean :2.331	Mean :2.961	Mean :3.922	Mean :2.275	Mean :1.574
3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:2.000
Max. :9.000	Max. :9.000	Max. :5.000	Max. :8.000	Max. :8.000	Max. :3.000
NA's :2387	NA's :2387	NA's :3242	NA's :3229	NA's :3229	
diabetes_06	numfalls24_06	arthritis_06	selfrhealth_06	age_06	marcat_08
Min. :0.0000	Min. : 0.000	Min. :0.0000	Min. :1.000	Min. : 52.00	Min. :1.000
1st Qu.:0.0000	1st Qu.: 0.000	1st Qu.:0.0000	1st Qu.:2.000	1st Qu.: 62.00	1st Qu.:1.000
Median :0.0000	Median : 0.000	Median :1.0000	Median :3.000	Median : 69.00	Median :2.000
Mean :0.2129	Mean : 0.864	Mean :0.6252	Mean :2.918	Mean : 69.52	Mean :1.576
3rd Qu.:0.0000	3rd Qu.: 1.000	3rd Qu.:1.0000	3rd Qu.:4.000	3rd Qu.: 77.00	3rd Qu.:2.000
Max. :1.0000	Max. :50.000	Max. :1.0000	Max. :5.000	Max. :104.00	Max. :3.000
	NA's :3976				NA's :1172
diabetes_08	numfalls24_08	arthritis_08	selfrhealth_08	age_08	marcat_10
Min. :0.0000	Min. : 0.000	Min. :0.0000	Min. :1.000	Min. : 54.00	Min. :1.000
1st Qu.:0.0000	1st Qu.: 0.000	1st Qu.:0.0000	1st Qu.:2.000	1st Qu.: 63.00	1st Qu.:1.000
Median :0.0000	Median : 0.000	Median :1.0000	Median :3.000	Median : 70.00	Median :2.000
Mean :0.2331	Mean : 1.121	Mean :0.6462	Mean :2.971	Mean : 70.97	Mean :1.582
3rd Qu.:0.0000	3rd Qu.: 1.000	3rd Qu.:1.0000	3rd Qu.:4.000	3rd Qu.: 78.00	3rd Qu.:2.000
Max. :1.0000	Max. :50.000	Max. :1.0000	Max. :5.000	Max. :106.00	Max. :3.000
NA's :1226	NA's :4255	NA's :1228	NA's :1225	NA's :1215	NA's :2359
diabetes_10	numfalls24_10	arthritis_10	selfrhealth_10	age_10	marcat_12
Min. :0.0000	Min. : 0.000	Min. :0.0000	Min. :1.000	Min. : 56.00	Min. :1.000
1st Qu.:0.0000	1st Qu.: 0.000	1st Qu.:0.0000	1st Qu.:2.000	1st Qu.: 65.00	1st Qu.:1.000
Median :0.0000	Median : 0.000	Median :1.0000	Median :3.000	Median : 72.00	Median :2.000
Mean :0.2586	Mean : 1.232	Mean :0.6675	Mean :2.926	Mean : 72.37	Mean :1.593
3rd Qu.:1.0000	3rd Qu.: 1.000	3rd Qu.:1.0000	3rd Qu.:4.000	3rd Qu.: 79.00	3rd Qu.:2.000
Max. :1.0000	Max. :50.000	Max. :1.0000	Max. :5.000	Max. :109.00	Max. :3.000
NA's :2399	NA's :4818	NA's :2409	NA's :2395	NA's :2387	NA's :3232
diabetes_12	numfalls24_12	arthritis_12	selfrhealth_12	age_12	edcat
Min. :0.000	Min. : 0.000	Min. :0.00	Min. :1.000	Min. : 58.00	Min. :1.000
1st Qu.:0.000	1st Qu.: 0.000	1st Qu.:0.00	1st Qu.:2.000	1st Qu.: 66.00	1st Qu.:2.000
Median :0.000	Median : 0.000	Median :1.00	Median :3.000	Median : 73.00	Median :2.000
Mean :0.268	Mean : 1.206	Mean :0.68	Mean :2.961	Mean : 73.58	Mean :2.396
3rd Qu.:1.000	3rd Qu.: 1.000	3rd Qu.:1.00	3rd Qu.:4.000	3rd Qu.: 80.00	3rd Qu.:3.000
Max. :1.000	Max. :50.000	Max. :1.00	Max. :5.000	Max. :103.00	Max. :4.000
NA's :3235	NA's :5082	NA's :3243	NA's :3242	NA's :3229	
racecat	ln_inc06	ln_inc08	ln_inc10	ln_inc12	kwgtr_dec
Min. :1.000	Min. : 0.000	Min. : 0.000	Min. : 0.00	Min. : 0.000	Min. :0.000
1st Qu.:2.000	1st Qu.: 9.647	1st Qu.: 9.686	1st Qu.: 9.70	1st Qu.: 9.627	1st Qu.:2.000
Median :2.000	Median :10.348	Median :10.388	Median :10.36	Median :10.242	Median :4.000
Mean :2.114	Mean :10.291	Mean :10.327	Mean :10.26	Mean :10.172	Mean :4.485
3rd Qu.:2.000	3rd Qu.:11.012	3rd Qu.:11.062	3rd Qu.:11.01	3rd Qu.:10.891	3rd Qu.:7.000

```

Max. :4.000 Max. :17.049 Max. :17.910 Max. :14.92 Max. :15.115 Max. :9.000
NA's :1215 NA's :2387 NA's :3229

```

```
>
```

```

> b$selfrhealth_06 <- factor(b$selfrhealth_06)
> b$marcat_06 <- factor(b$marcat_06)
> b$racecat <- factor(b$racecat)
> b$edcat <- factor(b$edcat)
> b$stratum <- factor(b$stratum)
> b$kwgtr_dec <- as.numeric(b$kwgtr_dec)
> # summary(b)

```

```

> # subset variables by number position in data set
> (hrs_mi_1wave_sub <- b[, c(10,11,13,41,42,44,45,46,65,66,67,68,71)])
# A tibble: 11,789 × 13

```

	secu	stratum	kwgtr	marcat_06	diabetes_06	arthritis_06	selfrhealth_06	age_06	edcat	racecat	ln_inc06
	<dbl>	<fct>	<dbl>	<fct>	<dbl>	<dbl>	<fct>	<dbl>	<fct>	<fct>	<dbl>
1	1	40	4093	1	0	0	3	70	2	2	10.6
2	2	1	7434	3	0	0	4	66	2	2	9.23
3	2	1	5217	1	0	1	4	66	4	2	11.3
4	2	1	5373	2	1	0	3	68	2	2	9.57
5	2	1	5440	2	0	0	2	58	3	2	10.9
6	2	2	5217	1	0	1	2	70	4	2	12.2
7	2	2	5778	2	0	0	2	64	4	4	10.9
8	2	2	5400	1	0	0	2	78	4	2	13.0
9	2	1	1799	2	1	0	4	68	1	1	9.24
10	2	1	3282	1	1	1	4	69	1	1	10.1

```

# i 11,779 more rows
# i 2 more variables: ln_inc08 <dbl>, kwgtr_dec <dbl>
# i Use `print(n = ...)` to see more rows

```

```

> summary(hrs_mi_1wave_sub)

```

	secu	stratum	kwgtr	marcat_06	diabetes_06	arthritis_06
Min.	:1.000	46 : 487	Min. : 924	1:5502	Min. :0.0000	Min. :0.0000
1st Qu.:	1.000	33 : 423	1st Qu.: 2433	2:5802	1st Qu.:0.0000	1st Qu.:0.0000
Median:	2.000	45 : 416	Median : 3723	3: 485	Median :0.0000	Median :1.0000
Mean	:1.503	40 : 380	Mean : 4458		Mean :0.2129	Mean :0.6252
3rd Qu.:	2.000	47 : 379	3rd Qu.: 5296		3rd Qu.:0.0000	3rd Qu.:1.0000
Max.	:2.000	29 : 357	Max. :17035		Max. :1.0000	Max. :1.0000

(Other):9347

```

selfrhealth_06 age_06 edcat racecat ln_inc06 ln_inc08 kwgtr_dec
1:1211 Min. : 52.00 1:2889 1:1039 Min. : 0.000 Min. : 0.000 Min. :0.000
2:3285 1st Qu.: 62.00 2:3903 2:8627 1st Qu.: 9.647 1st Qu.: 9.686 1st Qu.:2.000
3:3610 Median : 69.00 3:2438 3:1863 Median :10.348 Median :10.388 Median :4.000
4:2626 Mean : 69.52 4:2559 4: 260 Mean :10.291 Mean :10.327 Mean :4.485
5:1057 3rd Qu.: 77.00 3rd Qu.:11.012 3rd Qu.:11.062 3rd Qu.:7.000
Max. :104.00 Max. :17.049 Max. :17.910 Max. :9.000

```

```
> NA's :1215
```

```

> # use mice to examine and impute missing data
> md.pattern(hrs_mi_1wave_sub)
      secu stratum kwgtr marcat_06 diabetes_06 arthritis_06 selfrhealth_06 age_06 edcat racecat
10574  1      1      1          1           1           1           1      1      1      1
1215   1      1      1          1           1           1           1      1      1      1
      0      0      0          0           0           0           0      0      0      0
      ln_inc06 kwgtr_dec ln_inc08
10574      1          1          1      0
1215      1          1          0      1
      0          0      1215 1215
>
> ini <- mice(hrs_mi_1wave_sub, maxiter=0)

```

```

iter imp variable
  1  1  ln_inc08
  1  2  ln_inc08
  1  3  ln_inc08
  1  4  ln_inc08
  1  5  ln_inc08
  2  1  ln_inc08
  2  2  ln_inc08
  2  3  ln_inc08
  2  4  ln_inc08
  2  5  ln_inc08
  3  1  ln_inc08
  3  2  ln_inc08
  3  3  ln_inc08
  3  4  ln_inc08
  3  5  ln_inc08
  4  1  ln_inc08
  4  2  ln_inc08
  4  3  ln_inc08
  4  4  ln_inc08
  4  5  ln_inc08
  5  1  ln_inc08
  5  2  ln_inc08
  5  3  ln_inc08
  5  4  ln_inc08
  5  5  ln_inc08

```

```

> summary(ini)
Class: mids
Number of multiple imputations: 5
Imputation methods:

```

secu	stratum	kwgtr	marcat_06	diabetes_06	arthritis_06
""	""	""	""	""	""
selfrhealth_06	age_06	edcat	racecat	ln_inc06	ln_inc08
""	""	""	""	""	"pmm"
kwgtr_dec					
""					

PredictorMatrix:

	secu	stratum	kwgtr	marcat_06	diabetes_06	arthritis_06	selfrhealth_06	age_06	edcat
secu	0	1	1	1	1	1	1	1	1
stratum	1	0	1	1	1	1	1	1	1
kwgtr	1	1	0	1	1	1	1	1	1
marcat_06	1	1	1	0	1	1	1	1	1
diabetes_06	1	1	1	1	0	1	1	1	1
arthritis_06	1	1	1	1	1	0	1	1	1
	racecat	ln_inc06	ln_inc08	kwgtr_dec					
secu	1	1	1	1					

```

stratum      1      1      1      1
kwtgr       1      1      1      1
marcat_06   1      1      1      1
diabetes_06 1      1      1      1
arthritis_06 1     1      1      1
>
> # add a predictor matrix to control imputation model predictors for each imputed variable
> pred <- ini$predictorMatrix
> pred[,"kwtgr"] <- 0
> pred[,"secu"] <- 0
> pred[,"kwtgr_dec"] <- 1
> pred
      secu stratum kwtgr marcat_06 diabetes_06 arthritis_06 selfrhealth_06 age_06 edcat
secu      0      1      0      1      1      1      1      1      1
stratum   0      0      0      1      1      1      1      1      1
kwtgr     0      1      0      1      1      1      1      1      1
marcat_06 0      1      0      0      1      1      1      1      1
diabetes_06 0     1      0      1      0      1      1      1      1
arthritis_06 0     1      0      1      1      0      1      1      1
selfrhealth_06 0     1      0      1      1      1      0      1      1
age_06    0      1      0      1      1      1      1      0      1
edcat     0      1      0      1      1      1      1      1      0
racecat   0      1      0      1      1      1      1      1      1
ln_inc06  0      1      0      1      1      1      1      1      1
ln_inc08  0      1      0      1      1      1      1      1      1
kwtgr_dec 0      1      0      1      1      1      1      1      1
      racecat ln_inc06 ln_inc08 kwtgr_dec
secu      1      1      1      1
stratum   1      1      1      1
kwtgr     1      1      1      1
marcat_06 1      1      1      1
diabetes_06 1     1      1      1
arthritis_06 1     1      1      1
selfrhealth_06 1     1      1      1
age_06    1      1      1      1
edcat     1      1      1      1
racecat   0      1      1      1
ln_inc06  1      0      1      1
ln_inc08  1      1      0      1
kwtgr_dec 1      1      1      1
>
>
> imphrs1wave <- mice(hrs_mi_1wave_sub, m=5, pred=pred, seed=41279)

iter imp variable
 1  1  ln_inc08
 1  2  ln_inc08
 1  3  ln_inc08
 1  4  ln_inc08
 1  5  ln_inc08
 2  1  ln_inc08
 2  2  ln_inc08
 2  3  ln_inc08
 2  4  ln_inc08
 2  5  ln_inc08
 3  1  ln_inc08
 3  2  ln_inc08
 3  3  ln_inc08
 3  4  ln_inc08

```

```

3 5 ln_inc08
4 1 ln_inc08
4 2 ln_inc08
4 3 ln_inc08
4 4 ln_inc08
4 5 ln_inc08
5 1 ln_inc08
5 2 ln_inc08
5 3 ln_inc08
5 4 ln_inc08
5 5 ln_inc08

```

```
> (imphrs1wave)
```

```
Class: mids
```

```
Number of multiple imputations: 5
```

```
Imputation methods:
```

```

      secu      stratum      kwgtr      marcat_06      diabetes_06      arthritis_06
      ""         ""         ""         ""         ""         ""
selfrhealth_06      age_06      edcat      racecat      ln_inc06      ln_inc08
      ""         ""         ""         ""         ""         "pmm"
      kwgtr_dec
      ""

```

```
PredictorMatrix:
```

```

      secu stratum kwgtr marcat_06 diabetes_06 arthritis_06 selfrhealth_06 age_06 edcat
secu      0      1      0      1      1      1      1      1      1
stratum   0      0      0      1      1      1      1      1      1
kwgtr     0      1      0      1      1      1      1      1      1
marcat_06 0      1      0      0      1      1      1      1      1
diabetes_06 0      1      0      1      0      1      1      1      1
arthritis_06 0      1      0      1      1      0      1      1      1
      racecat ln_inc06 ln_inc08 kwgtr_dec
secu      1      1      1      1
stratum   1      1      1      1
kwgtr     1      1      1      1
marcat_06 1      1      1      1
diabetes_06 1      1      1      1
arthritis_06 1      1      1      1

```

```
> summary(imphrs1wave)
```

```
Class: mids
```

```
Number of multiple imputations: 5
```

```
Imputation methods:
```

```

      secu      stratum      kwgtr      marcat_06      diabetes_06      arthritis_06
      ""         ""         ""         ""         ""         ""
selfrhealth_06      age_06      edcat      racecat      ln_inc06      ln_inc08
      ""         ""         ""         ""         ""         "pmm"
      kwgtr_dec
      ""

```

```
PredictorMatrix:
```

```

      secu stratum kwgtr marcat_06 diabetes_06 arthritis_06 selfrhealth_06 age_06 edcat
secu      0      1      0      1      1      1      1      1      1
stratum   0      0      0      1      1      1      1      1      1
kwgtr     0      1      0      1      1      1      1      1      1
marcat_06 0      1      0      0      1      1      1      1      1
diabetes_06 0      1      0      1      0      1      1      1      1
arthritis_06 0      1      0      1      1      0      1      1      1
      racecat ln_inc06 ln_inc08 kwgtr_dec
secu      1      1      1      1

```

```

stratum      1      1      1      1
kwtgr        1      1      1      1
marcat_06    1      1      1      1
diabetes_06  1      1      1      1
arthritis_06 1      1      1      1

```

```
> # Combine all five data.frame objects into a list
```

```
> list_of_data_frames <-
+ list(
+ mice::complete( imphrs1wave , 1 ) ,
+ mice::complete( imphrs1wave , 2 ) ,
+ mice::complete( imphrs1wave , 3 ) ,
+ mice::complete( imphrs1wave , 4 ) ,
+ mice::complete( imphrs1wave , 5 )
+ )
```

```
> # Coerce weight variable to numeric (coercing directly factor to numeric here changes the levels)
```

```
> list_of_data_frames <-
+ lapply(list_of_data_frames ,
+ function( w ) transform( w , kwgtr_dec = as.numeric( as.character( kwgtr_dec ) ) )
+ )
```

```
> # multiply-imputed design statement
```

```
> deshrs_1wave <-
+ svydesign(
+ id = ~ secu ,
+ strat = ~ stratum ,
+ weight = ~ kwgtr_dec ,
+ data = imputationList(list_of_data_frames ) ,
+ nest = TRUE
+ )
>
```

```
> # Multiply-imputed average & SE of that variable
```

```
> MIcombine(with(deshrs_1wave , svymean( ~ ln_inc08 ) ) )
```

```
Multiple imputation results:
```

```
      with(deshrs_1wave, svymean(~ln_inc08))
      MIcombine.default(with(deshrs_1wave, svymean(~ln_inc08)))
      results      se
ln_inc08 10.41422 0.02577145
```

```
>
```

```
> summary(m1w)
```

```
Error in h(simpleError(msg, call)) :
```

```
error in evaluating the argument 'object' in selecting a method for function 'summary': object 'm1w' not found
```

```
> exp(10.41422) #mean
```

```
[1] 33330.23
```

```
> exp(10.36371) #lci
```

```
[1] 31688.53
```

```
> exp(10.46473) #uci
```

```
[1] 35056.98
```

```
>
```

```
>
```

```
> # multiple Imputation using a Selection Model Not Available in R
```

```
>
```

```
>
```

```

> ## 2 Waves of Data
>
> # Complete Case 2 Waves
> hrs_2wave <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/cc_2waves.sas7bdat")
> names(hrs_2wave)
 [1] "hhid"          "pn"            "kfinr"         "KC001"         "KC010"
 [6] "KC070"         "gender"        "hispanic"      "schlyrs"       "secu"
[11] "stratum"       "kmarst"        "kwgtr"         "KWHYORWT"     "lfinr"
[16] "lmarst"        "lwgtr"         "mfinr"         "mmarst"        "mwgtr"
[21] "nfinr"         "nmarst"        "nwgtr"         "H8ATOTA"       "H9ATOTA"
[26] "H10ATOTA"     "H11ATOTA"     "H8ITOT"        "H9ITOT"        "H10ITOT"
[31] "H11ITOT"     "LC001"         "LC010"         "LC070"         "MC001"
[36] "MC010"         "MC070"         "NC001"         "NC010"         "NC070"
[41] "marcat_06"     "diabetes_06"   "numfalls24_06" "arthritis_06"  "selfrhealth_06"
[46] "age_06"        "marcat_08"     "diabetes_08"   "numfalls24_08" "arthritis_08"
[51] "selfrhealth_08" "age_08"        "marcat_10"     "diabetes_10"   "numfalls24_10"
[56] "arthritis_10"  "selfrhealth_10" "age_10"        "marcat_12"     "diabetes_12"
[61] "numfalls24_12" "arthritis_12"  "selfrhealth_12" "age_12"        "edcat"
[66] "racecat"       "ln_inc06"      "ln_inc08"      "ln_inc10"      "ln_inc12"
[71] "incdiff_06_10" "resp10"
>
> svyhrs_cc_2 <- svydesign(strata=~stratum, id=~secu, weights=~kwgtr, data=hrs_2wave, nest=T)
>
> ex11_2 <- svymean(~incdiff_06_10, design=svyhrs_cc_2, se=T, ci=T, keep.vars=T, na.rm=T)
> show(ex11_2)
      mean      SE
incdiff_06_10 -6551.4 1866.1
> confint(ex11_2)
      2.5 %    97.5 %
incdiff_06_10 -10208.96 -2893.845

> # Adjusted Weight 2 Wave
> hrs_2waves_adj <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R
Code/adj_wgt_2waves.sas7bdat")
> names(hrs_2waves_adj)
 [1] "hhid"          "pn"            "kfinr"         "KC001"         "KC010"
 [6] "KC070"         "gender"        "hispanic"      "schlyrs"       "secu"
[11] "stratum"       "kmarst"        "kwgtr"         "KWHYORWT"     "lfinr"
[16] "lmarst"        "lwgtr"         "mfinr"         "mmarst"        "mwgtr"
[21] "nfinr"         "nmarst"        "nwgtr"         "H8ATOTA"       "H9ATOTA"
[26] "H10ATOTA"     "H11ATOTA"     "H8ITOT"        "H9ITOT"        "H10ITOT"
[31] "H11ITOT"     "LC001"         "LC010"         "LC070"         "MC001"
[36] "MC010"         "MC070"         "NC001"         "NC010"         "NC070"
[41] "marcat_06"     "diabetes_06"   "numfalls24_06" "arthritis_06"  "selfrhealth_06"
[46] "age_06"        "marcat_08"     "diabetes_08"   "numfalls24_08" "arthritis_08"
[51] "selfrhealth_08" "age_08"        "marcat_10"     "diabetes_10"   "numfalls24_10"
[56] "arthritis_10"  "selfrhealth_10" "age_10"        "marcat_12"     "diabetes_12"
[61] "numfalls24_12" "arthritis_12"  "selfrhealth_12" "age_12"        "edcat"
[66] "racecat"       "ln_inc06"      "ln_inc08"      "ln_inc10"      "ln_inc12"
[71] "incdiff_06_10" "resp10"        "_LEVEL_"       "phat1"         "dec"
[76] "mean_phat"     "adj_kwgtr"
>
> svyhrs_adj_1 <- svydesign(strata=~stratum, id=~secu, weights=~adj_kwgtr, data=hrs_2waves_adj, nest=T)
>
> show(ex11_2_adj <- svymean(~incdiff_06_10, design=svyhrs_adj_1, se=T, ci=T, keep.vars=T, na.rm=T))
      mean      SE
incdiff_06_10 -6120 1703

```

```

> confint(ex11_2_adj)
                2.5 %   97.5 %
incdiff_06_10 -9457.72 -2782.22
>

> # Multiple Imputation for 2 Waves, SAS data set already prepared for this example
> hrs_a <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/wt_deciles_2.sas7bdat")
> names(hrs_a)
 [1] "HHID"          "PN"            "KFINR"         "KC001"         "KC010"
 [6] "KC070"         "GENDER"        "HISPANIC"      "SCHLYRS"       "SECU"
[11] "STRATUM"       "KMARST"        "KWGTR"         "KWHYORWT"     "LFINR"
[16] "LMARST"        "LWGTR"         "MFINR"         "MMARST"        "MWGTR"
[21] "NFINR"         "NMARST"        "NWGTR"         "H8ATOTA"       "H9ATOTA"
[26] "H10ATOTA"     "H11ATOTA"     "H8ITOT"        "H9ITOT"        "H10ITOT"
[31] "H11ITOT"      "LC001"         "LC010"         "LC070"         "MC001"
[36] "MC010"        "MC070"         "NC001"         "NC010"         "NC070"
[41] "marcat_06"     "diabetes_06"   "numfalls24_06" "arthritis_06"  "selfrhealth_06"
[46] "age_06"        "marcat_08"     "diabetes_08"   "numfalls24_08" "arthritis_08"
[51] "selfrhealth_08" "age_08"        "marcat_10"     "diabetes_10"   "numfalls24_10"
[56] "arthritis_10"  "selfrhealth_10" "age_10"        "marcat_12"     "diabetes_12"
[61] "numfalls24_12" "arthritis_12"  "selfrhealth_12" "age_12"        "edcat"
[66] "racecat"       "ln_inc06"      "ln_inc08"      "ln_inc10"      "ln_inc12"
[71] "kwgtr_dec"     "incdiff_06_10"
>
> hrs_a$selfrhealth_06 <- factor(hrs_a$selfrhealth_06)
> hrs_a$marcat_06 <- factor(hrs_a$marcat_06)
> hrs_a$racecat <- factor(hrs_a$racecat)
> hrs_a$edcat <- factor(hrs_a$edcat)
> hrs_a$STRATUM <- factor(hrs_a$STRATUM)
> hrs_a$kwgtr_dec <- factor(hrs_a$kwgtr_dec)
>
> # Subset variables by number position in data set, keep those needed for imputation
> hrs_mi_2waves_sub <- hrs_a [, c(10,11,13,41,42,44,45,46,65,66,67,69,71)]
> names(hrs_mi_2waves_sub)
 [1] "SECU"          "STRATUM"       "KWGTR"         "marcat_06"     "diabetes_06"
 [6] "arthritis_06"  "selfrhealth_06" "age_06"        "edcat"         "racecat"
[11] "ln_inc06"      "ln_inc10"      "kwgtr_dec"
>
> # Use mice to impute missing data
> library(mice)
>
> # Dry run to prepare the predictor matrix
> ini <- mice(hrs_mi_2waves_sub, maxiter=0)

```

```

iter imp variable
 1  1  ln_inc10
 1  2  ln_inc10
 1  3  ln_inc10
 1  4  ln_inc10
 1  5  ln_inc10
 2  1  ln_inc10
 2  2  ln_inc10
 2  3  ln_inc10
 2  4  ln_inc10
 2  5  ln_inc10
 3  1  ln_inc10
 3  2  ln_inc10
 3  3  ln_inc10

```

```

3 4 ln_inc10
3 5 ln_inc10
4 1 ln_inc10
4 2 ln_inc10
4 3 ln_inc10
4 4 ln_inc10
4 5 ln_inc10
5 1 ln_inc10
5 2 ln_inc10
5 3 ln_inc10
5 4 ln_inc10
5 5 ln_inc10

```

```
> summary(ini)
```

```
Class: mids
```

```
Number of multiple imputations: 5
```

```
Imputation methods:
```

```

          SECU          STRATUM          KWGTR          marcat_06          diabetes_06          arthritis_06
          ""           ""           ""           ""           ""           ""
selfrhealth_06          age_06          edcat          racecat          ln_inc06          ln_inc10
          ""           ""           ""           ""           ""           "pmm"
          kwgtr_dec
          ""

```

```
PredictorMatrix:
```

```

          SECU STRATUM KWGTR marcat_06 diabetes_06 arthritis_06 selfrhealth_06 age_06 edcat
SECU      0      1      1          1          1          1          1          1      1
STRATUM   1      0      1          1          1          1          1          1      1
KWGTR     1      1      0          1          1          1          1          1      1
marcat_06 1      1      1          0          1          1          1          1      1
diabetes_06 1      1      1          1          0          1          1          1      1
arthritis_06 1      1      1          1          1          0          1          1      1

          racecat ln_inc06 ln_inc10 kwgtr_dec
SECU      1      1          1          1
STRATUM   1      1          1          1
KWGTR     1      1          1          1
marcat_06 1      1          1          1
diabetes_06 1      1          1          1
arthritis_06 1      1          1          1

```

```

>
> # Add a predictor matrix to control imputation model predictors for each imputed variable
> pred <- ini$predictorMatrix
> pred[,"KWGTR"] <- 0
> pred[,"SECU"] <- 0
> pred[,"kwgtr_dec"] <- 1
> pred[,"ln_inc06"] <- 1
> pred

```

```

          SECU STRATUM KWGTR marcat_06 diabetes_06 arthritis_06 selfrhealth_06 age_06 edcat
SECU      0      1      0          1          1          1          1          1      1
STRATUM   0      0      0          1          1          1          1          1      1
KWGTR     0      1      0          1          1          1          1          1      1
marcat_06 0      1      0          0          1          1          1          1      1
diabetes_06 0      1      0          1          0          1          1          1      1
arthritis_06 0      1      0          1          1          0          1          1      1
selfrhealth_06 0      1      0          1          1          1          0          1      1
age_06    0      1      0          1          1          1          1          0      1
edcat     0      1      0          1          1          1          1          1      0
racecat   0      1      0          1          1          1          1          1      1
ln_inc06  0      1      0          1          1          1          1          1      1
ln_inc10  0      1      0          1          1          1          1          1      1
kwgtr_dec 0      1      0          1          1          1          1          1      1

```

```

      racecat ln_inc06 ln_inc10 kwgtr_dec
SECU          1          1          1          1
STRATUM       1          1          1          1
KWGTR         1          1          1          1
marcat_06     1          1          1          1
diabetes_06   1          1          1          1
arthritis_06  1          1          1          1
selfrhealth_06 1          1          1          1
age_06        1          1          1          1
edcat         1          1          1          1
racecat       0          1          1          1
ln_inc06      1          1          1          1
ln_inc10      1          1          0          1
kwgtr_dec     1          1          1          1
>
>
> # Use same variables as from C11 Stata example
> imphrs2waves1 <- mice(hrs_mi_2waves_sub, pred=pred, m=5, seed=41279, method="norm.nob", print=FALSE)
> imphrs2waves1
Class: mids
Number of multiple imputations: 5
Imputation methods:
      SECU          STRATUM          KWGTR          marcat_06          diabetes_06          arthritis_06
      ""           ""           ""           ""           ""           ""
selfrhealth_06          age_06          edcat          racecat          ln_inc06          ln_inc10
      ""           ""           ""           ""           ""           ""
      kwgtr_dec
      ""
PredictorMatrix:
      SECU STRATUM KWGTR marcat_06 diabetes_06 arthritis_06 selfrhealth_06 age_06 edcat
SECU          0          1          0          1          1          1          1          1          1
STRATUM       0          0          0          1          1          1          1          1          1
KWGTR         0          1          0          1          1          1          1          1          1
marcat_06     0          1          0          0          1          1          1          1          1
diabetes_06   0          1          0          1          0          1          1          1          1
arthritis_06  0          1          0          1          1          0          1          1          1
      racecat ln_inc06 ln_inc10 kwgtr_dec
SECU          1          1          1          1
STRATUM       1          1          1          1
KWGTR         1          1          1          1
marcat_06     1          1          1          1
diabetes_06   1          1          1          1
arthritis_06  1          1          1          1
>
>
> # Combine all five data.frame objects into a list
> list_of_data_frames <-
+ list(
+ mice::complete( imphrs2waves1 , 1 ) ,
+ mice::complete( imphrs2waves1 , 2 ) ,
+ mice::complete( imphrs2waves1 , 3 ) ,
+ mice::complete( imphrs2waves1 , 4 ) ,
+ mice::complete( imphrs2waves1 , 5 )
+ )

```

```

> # Coerce weight variable to numeric
> list_of_data_frames <-
+ lapply(
+ list_of_data_frames ,
+ function( w ) transform( w , kwgtr_dec = as.numeric( as.character( kwgtr_dec ) ) )
+ )

> # Set survey design , use 2006 individual weight for survey design setup
> # Multiply-imputed design statement
> deshrs_2waves <-
+ svydesign(
+ id = ~ SECU ,
+ strat = ~ STRATUM ,
+ weight = ~ KWGTR ,
+ data = imputationList( list_of_data_frames ) ,
+ nest = TRUE)
> deshrs_2waves
Multiple (5) imputations: svydesign(id = ~SECU, strat = ~STRATUM, weight = ~KWGTR, data =
imputationList(list_of_data_frames),
  nest = TRUE)
>
> # Control values used in analysis
> deshrs_2waves <- update(deshrs_2waves, ln_inc10a=ifelse(ln_inc10 > 14.92, 14.92, ln_inc10),
inc10=exp(ln_inc10a), inc06=exp(ln_inc06))
> deshrs_2waves <- update(deshrs_2waves, new_chg0610=(inc10 - inc06))
> deshrs_2waves <- update(deshrs_2waves, new_chg0610a=ifelse(new_chg0610 < -12300000, -12300000, new_chg0610),
+ new_chg0610b=ifelse(new_chg0610a > 2062968, 2062968, new_chg0610a))
> hrs_2w_meandiff <- with(deshrs_2waves, svymean(~(new_chg0610b), se=T, ci=T ))
> hrs_2w_meandiff
[[1]]
      mean      SE
new_chg0610b -2857.1 2807.4

[[2]]
      mean      SE
new_chg0610b -3326.9 2966.1

[[3]]
      mean      SE
new_chg0610b -2590.5 3051.1

[[4]]
      mean      SE
new_chg0610b -3351.2 2898.9

[[5]]
      mean      SE
new_chg0610b -3437.1 2995.9

attr(,"call")
with(deshrs_2waves, svymean(~(new_chg0610b), se = T, ci = T))
>
>

```

```

> # Use MIcombine for overall combined and design-adjusted mean/se
> (hrs_2w_comb <- MIcombine(hrs_2w_meandiff))
Multiple imputation results:
  with(deshrs_2waves, svymean(~(new_chg0610b), se = T, ci = T))
  MIcombine.default(hrs_2w_meandiff)
      results      se
new_chg0610b -3112.552 2972.779
> (hrs_2w_comb)
Multiple imputation results:
  with(deshrs_2waves, svymean(~(new_chg0610b), se = T, ci = T))
  MIcombine.default(hrs_2w_meandiff)
      results      se
new_chg0610b -3112.552 2972.779
> confint(hrs_2w_comb)
      2.5 %    97.5 %
new_chg0610b -8939.092 2713.989
>
> # No Selection Model Multiple Imputation Available in R
>

> # Calibration Method
> hrs_2waves_cal <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R
Code/calibration_2waves.sas7bdat")
> names(hrs_2waves_cal)
 [1] "hhid"      "pn"        "kfinr"     "KC001"     "KC010"
 [6] "KC070"     "gender"    "hispanic"  "schlyrs"   "secu"
[11] "stratum"   "kmarst"    "kwgtr"     "KWHYORWT"  "lfinr"
[16] "lmarst"    "lwgtr"     "mfinr"     "mmarst"    "mwgtr"
[21] "nfinr"     "nmarst"    "nwgtr"     "H8ATOTA"   "H9ATOTA"
[26] "H10ATOTA"  "H11ATOTA"  "H8ITOT"    "H9ITOT"    "H10ITOT"
[31] "H11ITOT"   "LC001"     "LC010"     "LC070"     "MC001"
[36] "MC010"     "MC070"     "NC001"     "NC010"     "NC070"
[41] "marcat_06" "diabetes_06" "numfalls24_06" "arthritis_06" "selfrhealth_06"
[46] "age_06"    "marcat_08"  "diabetes_08" "numfalls24_08" "arthritis_08"
[51] "selfrhealth_08" "age_08"    "marcat_10"  "diabetes_10"  "numfalls24_10"
[56] "arthritis_10" "selfrhealth_10" "age_10"    "marcat_12"    "diabetes_12"
[61] "numfalls24_12" "arthritis_12" "selfrhealth_12" "age_12"      "edcat"
[66] "racecat"   "ln_inc06"   "ln_inc10"   "_TYPE_"      "_FREQ_"
[71] "popsize"   "sumrespwgts" "cal_adj"    "resp10"     "kwgtr_cal"
[76] "incdiff_06_10"
>
> # Subset of cases without missing weight variable
> hrs_2waves_calsub <- hrs_2waves_cal[ which(hrs_2waves_cal$kwgtr_cal > 0),]
> summary(hrs_2waves_calsub$kwgtr_cal)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  970.3  3094.8  4610.8  5589.9  6814.2 23512.1
>
> # Get mean diff 06-10
> svyhrs_cal_2 <- svydesign(strata=~stratum, id=~secu, weights=~kwgtr_cal, data=hrs_2waves_calsub, nest=T,
na.rm=T)
> show(ex11_2_cal <- svymean(~incdiff_06_10, design=svyhrs_cal_2, se=T, ci=T, keep.vars=T, na.rm=T))
      mean      SE
incdiff_06_10 -6341.7 1780.6
> confint(ex11_2_cal)
      2.5 %    97.5 %
incdiff_06_10 -9831.568 -2851.746

```

```

> ## 3+ Waves of Data
> # Example 11.3.3 Weighted Multilevel Modeling not available in R Survey Package
> # Example 11.3.3.1 Veiga Method for multi-level modeling not available in R Survey Package
> # Example 11.3.4 Weighted GEE Analysis using Geepack from R ( See geepack.pdf for details)

> # Data set prepared using SAS data management and estimation tools, not repeated here, see either Stata or SAS
code for details
> # Load geepack
> library(geepack)
Warning message:
package 'geepack' was built under R version 4.4.3
> hrs_3w_gee <- read_sas("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/wgt_gee_3pwaves.sas7bdat")
> names(hrs_3w_gee)
 [1] "hhid"          "pn"            "gender"        "secu"          "stratum"       "marcat_06"
 [7] "diabetes_06"  "arthritis_06" "edcat"         "racecat"       "cumprob_case" "ln_inc"
[13] "year"         "basewgt"       "casewt"        "yrssince06"   "yrs06sq"       "newid"
[19] "newid_num"

> # set factor variables
> hrs_3w_gee$gender <- as.factor(hrs_3w_gee$gender)
> hrs_3w_gee$stratum <- as.factor(hrs_3w_gee$stratum)
> hrs_3w_gee$year <- as.factor(hrs_3w_gee$year)

> # Example 11.3.4
> # model formula
> mf <- formula(ln_inc~yrssince06 + gender + yrs06sq + (yrssince06*gender) + (yrs06sq*gender) + stratum)
> mf
ln_inc ~ yrssince06 + gender + yrs06sq + (yrssince06 * gender) +
  (yrs06sq * gender) + stratum
>
> # Run model using geeglm with weight
> ex11_3_4 <- geeglm(mf, data=hrs_3w_gee, id=newid_num, weight=casewt, family=gaussian("identity"),
corstr="exchangeable")
> summary(ex11_3_4)

```

```

Call:
geeglm(formula = mf, family = gaussian("identity"), data = hrs_3w_gee,
weights = casewt, id = newid_num, corstr = "exchangeable")

```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W )	
(Intercept)	9.811754	0.439507	498.382	< 2e-16	***
yrssince06	-0.083205	0.046595	3.189	0.07415	.
gender2	-0.633484	0.102824	37.956	7.24e-10	***
yrs06sq	0.004962	0.008031	0.382	0.53666	
stratum2	0.252472	0.495100	0.260	0.61009	
stratum3	0.756543	0.490653	2.377	0.12310	
stratum4	-0.294685	1.189041	0.061	0.80426	
stratum5	0.674029	1.666011	0.164	0.68579	
stratum6	0.902298	0.615887	2.146	0.14291	
stratum7	1.477010	0.512809	8.296	0.00397	**
stratum8	0.932406	0.480551	3.765	0.05234	.
stratum9	0.716107	0.520425	1.893	0.16882	
stratum10	1.484322	0.459878	10.418	0.00125	**
stratum11	0.238604	0.786072	0.092	0.76148	
stratum12	1.194361	0.545861	4.787	0.02867	*
stratum13	1.043130	0.524263	3.959	0.04662	*

stratum14	0.981955	0.471417	4.339	0.03725	*
stratum15	0.805363	0.511629	2.478	0.11546	
stratum16	0.837765	0.455406	3.384	0.06583	.
stratum17	1.287055	0.483077	7.098	0.00772	**
stratum18	1.397770	0.492931	8.041	0.00457	**
stratum19	1.072590	0.660809	2.635	0.10456	
stratum20	1.088616	0.738265	2.174	0.14033	
stratum21	1.171977	0.453173	6.688	0.00971	**
stratum22	1.120087	0.491161	5.201	0.02258	*
stratum23	0.557142	0.513011	1.179	0.27747	
stratum24	0.959569	0.527894	3.304	0.06911	.
stratum25	0.924567	0.454931	4.130	0.04212	*
stratum26	1.244131	0.535822	5.391	0.02024	*
stratum27	1.227339	0.465830	6.942	0.00842	**
stratum28	0.999721	0.464745	4.627	0.03147	*
stratum29	1.256069	0.473722	7.030	0.00801	**
stratum30	0.972002	0.476475	4.162	0.04135	*
stratum31	1.139282	0.464244	6.022	0.01413	*
stratum32	1.007486	0.620234	2.639	0.10430	
stratum33	0.934112	0.571731	2.669	0.10229	
stratum34	0.391956	0.456093	0.739	0.39013	
stratum35	1.140609	0.513953	4.925	0.02647	*
stratum36	0.769334	0.617221	1.554	0.21260	
stratum37	0.730089	0.453703	2.589	0.10758	
stratum38	1.357772	0.535490	6.429	0.01123	*
stratum39	0.961475	0.459745	4.374	0.03650	*
stratum40	1.418981	0.457088	9.637	0.00191	**
stratum41	1.322078	0.508997	6.747	0.00939	**
stratum42	0.673105	0.467798	2.070	0.15018	
stratum43	1.048939	0.478674	4.802	0.02843	*
stratum44	1.162143	0.464574	6.258	0.01237	*
stratum45	1.375144	0.458912	8.979	0.00273	**
stratum46	1.080036	0.475169	5.166	0.02303	*
stratum47	0.680106	0.453804	2.246	0.13396	
stratum48	0.938204	0.458681	4.184	0.04081	*
stratum49	0.352605	0.616467	0.327	0.56734	
stratum50	0.804207	0.457144	3.095	0.07854	.
stratum51	0.968136	0.485094	3.983	0.04596	*
stratum52	0.076949	0.502231	0.023	0.87823	
stratum53	0.983399	0.481331	4.174	0.04104	*
stratum54	0.292583	0.618121	0.224	0.63597	
stratum55	0.318202	0.518565	0.377	0.53947	
stratum56	1.342582	0.475870	7.960	0.00478	**
yrssince06:gender2	0.092424	0.077052	1.439	0.23033	
gender2:yrss06sq	-0.009592	0.012510	0.588	0.44323	
---					

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	2.286	0.2756

Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha	0.3487	0.09045

Number of clusters: 11789 Maximum cluster size: 4