

ASDA3 Analysis Example Replication Chapter 12

R Code

```
# Load packages needed
library(survey)
library(mice)
library(haven)
library(mitools)

# Read in Nhanes subset, set up for missing data imputation and analysis, data management done in SAS

nhanesc12 <- read_sas("P:/ASDA3/Replication SAS/c12_nhanes_subset.sas7bdat")
nhanesc12_sub <- nhanesc12 # make sure to note this is a subset n=5615

str(nhanesc12_sub)
md.pattern(nhanesc12_sub)

# Set survey design
nhanessvy <- svydesign(strata=~sdmvstra, id=~sdmvpsu, weights=~WTMEC2YR, data=nhanesc12_sub, nest=T)

# Complete Case Analysis
# Obtain means for 3 continuous variables without imputation, Table 12.3
print(ex12_3a <- svymean(~bmxmbmi, nhanessvy, se=T, na.rm=T, ci=T))
show(ex12_3b <- svymean(~BPXDI1, nhanessvy, se=T, na.rm=T, ci=T))
show(ex12_3c <- svymean(~indfmpir, nhanessvy, se=T, na.rm=T, ci=T))

# High Blood Pressure Mean, Complete Case, Table 12.4
(ex12_4 <- svymean(~factor(high_diastolic),nhanessvy, se=T, na.rm=T, deff=T, ci=T, keep.vars=T))
confint(ex12_4)

# Logistic Regression Complete Case Analysis with Design Correction, Table 12.5
mod12_5 <- svyglm(high_diastolic ~ factor(RIDRETH1) + factor(riagendr)+ agec + agecsq,family=quasibinomial,
design=nhanessvy)
summary(mod12_5)

# Imputation Method 1 with Design Variables in MI Model, return to original data without the CC high diastolic
blood pressure

nhanesc12_subm1 <- nhanesc12
summary(nhanesc12_subm1)
names(nhanesc12_subm1)

# Subset variables by number position in data set , drop age and high_diastolic since will be redone once
bpxdi1 is complete
(nhanesc12_subset <- nhanesc12[, c(1,2,3,4,5,6,7,8,10,12,13,14)])
summary(nhanesc12_subset)

# Use mice to impute and specify type of default method for imputation models
# Run without custom predictor matrix first,see which variables might not be used in final model

imp1 <- mice(nhanesc12_subset, n.imp=5, seed=2024, defaultMethod=c("norm","logreg","polyreg"))
imp1$predictorMatrix
summary(imp1)

# Add a predictor matrix to control imputation model predictors for each imputed variable
pred <- imp1$predictorMatrix
pred[,"sdmvpsu"] <- 0
```

```

pred[, "sdmvstra"] <- 0
pred[, "descode"] <- 1
pred[, "WTMEC2YR"] <- 1

pred #note that certain variables are set to 0 for this matrix

imp1_pred <- mice(nhanesc12_subset, pred=pred, n.imp=5, seed=2016, defaultMethod=c("norm", "logreg", "polyreg"))
summary(imp1_pred)

# Note small differences between 1st and 2nd impute
summary(mice::complete( imp1_pred , 1 )$BPXDI1)
summary(mice::complete( imp1_pred , 2 )$BPXDI1)

# Combine all five data.frame objects into a list
list_of_data_frames <- list(
mice::complete(imp1_pred , 1 ) ,
mice::complete( imp1_pred , 2 ) ,
mice::complete( imp1_pred , 3 ) ,
mice::complete( imp1_pred , 4 ) ,
mice::complete( imp1_pred , 5 ))

# Multiply-imputed design statement
desnhanesm1 <- svydesign(id = ~ sdmvpsu, strat = ~ sdmvstra , weight = ~ WTMEC2YR ,
  data = imputationList( list_of_data_frames ) , nest = TRUE )
desnhanesm1

# Add a new variable to desnhanesm1 using imputed bpxdi1 values
desnhanesm1 <- update(desnhanesm1, high_diastolic = ifelse(BPXDI1 >=90,1,0) )
(desnhanesm1)

# Combined mean high blood pressure with design adjustment, Table 12.4
fitm1_mean <- MIcombine(with(desnhanesm1, svymean(~factor(high_diastolic), se=T, ci=T, na.rm=T)))
fitm1_mean
confint(fitm1_mean)

# Run design based logistic model with svyglm using 5 imputed data sets contained in desnhanesm1, Table 12.5
fitm1 <- MIcombine(with(desnhanesm1, svyglm (factor(high_diastolic) ~ factor(RIDRETH1) + factor(riagendr) +
agec + agecsq, family=quasibinomial)))
fitm1

# Obtain means for 3 continuous variables imputed using method 1, Table 12.3 (by each replicate)
fitm1_ex12_3 <- with(desnhanesm1, svymean(~bmx bmi + BPXDI1 + indfmpir), se=T, na.rm=T, ci=T )
fitm1_ex12_3

# Imputation 2 NO design variables in models
# Return to original data without the CC high diastolic blood pressure

nhanesc12_submeth2 <- nhanesc12
names(nhanesc12_submeth2)

# Subset variables by number position in data set , drop high_diastolic for imputations
(nhanesc12_subset <- nhanesc12[, c(1,2,3,4,5,6,7,8,10, 12,13,14)])
summary(nhanesc12_subset)

nhanesc12_subm2 <- nhanesc12_subset

# Add a predictor matrix to control imputation model predictors for each imputed variable
pred <- imp1_pred$predictorMatrix

```

```

pred[, "sdmvpsu"] <- 0
pred[, "sdmvstra"] <- 0
pred[, "descode"] <- 0
pred[, "WTMEC2YR"] <- 0
imp2$predictorMatrix

pred

imp2_pred <- mice(nhanesc12_subm2, pred=pred, n.imp=5, seed=2024, defaultMethod=c("norm", "logreg", "polyreg"))
summary(imp2_pred)

# Small differences between 1st and 2nd implicate
summary(mice::complete( imp2_pred , 1 )$BPXDI1)
summary(mice::complete( imp2_pred , 2 )$BPXDI1)

# Combine all five data.frame objects into a list
list_of_data_frames2 <- list(
mice::complete( imp2_pred , 1 ) ,
mice::complete( imp2_pred , 2 ) ,
mice::complete( imp2_pred , 3 ) ,
mice::complete( imp2_pred , 4 ) ,
mice::complete( imp2_pred , 5 )
)

# Multiply-imputed design statement
desnhanesm2 <- svydesign(id = ~ sdmvpsu, strat = ~ sdmvstra , weight = ~ WTMEC2YR ,
  data = imputationList( list_of_data_frames2 ), nest = TRUE)
desnhanesm2

# Add a new variable to desnhanesm2 using imputed blood pressure values
desnhanesm2 <- update( desnhanesm2, high_diastolic = ifelse(BPXDI1 >=90,1,0)
(desnhanesm2)

# Combined mean high blood pressure with NO design adjustment, Table 12.4
fitm2_mean <- MIcombine(with(desnhanesm2, svymean(~factor(high_diastolic), se=T, na.rm=T, ci=T )))
fitm2_mean

confint(fitm2_mean)

# Run design based logistic model with svyglm using 5 imputed data sets contained in desnhanesm2, Table 12.5
fitm2 <- MIcombine(with(desnhanesm2, svyglm (factor(high_diastolic) ~ factor(RIDRETH1) + factor(riagendr) +
agec + agecsq, family=quasibinomial)))
(fitm2)

# Obtain means for 3 continuous variables imputed using method 2 (NO DESIGN) , Table 12.3 (by each replicate)
fitm2_ex12_3 <- with(desnhanesm2, svymean(~bmx bmi + BPXDI1 + indfmpir, se=T, na.rm=T, ci=T ))
fitm2_ex12_3

# Note: FEFI method not available in R Survey Package, see examples using other software such as SAS PROC
SURVEYIMPUTE

```

R Results

```
> # Chapter 12 Multiple Imputation Using R mice and mitools with Survey Package
>
> # Load packages needed
> library(survey)
> library(mice)
> library(haven)
> library(mitools)
>
> # Read in Nhanes subset, set up for missing data imputation and analysis, data management done in SAS
>
> nhanesc12 <- read_sas("P:/ASDA3/Replication SAS/c12_nhanes_subset.sas7bdat")
> nhanesc12_sub <- nhanesc12 # make sure to note this is a subset n=5615
>
> str(nhanesc12_sub)
tibble [5,615 × 15] (S3: tbl_df/tbl/data.frame)
 $ riagendr      : num [1:5615] 1 2 1 2 1 2 1 1 1 1 ...
  .. attr(*, "label")= chr "Gender"
 $ RIDRETH1      : num [1:5615] 3 3 5 4 3 3 5 3 5 3 ...
  .. attr(*, "label")= chr "1=mex 2=oth hisp 3=white 4=black 5=other"
 $ WTMEC2YR      : num [1:5615] 104237 127965 14784 27123 27336 ...
  .. attr(*, "label")= chr "Full sample 2 year MEC exam weight"
 $ sdmvpsu       : num [1:5615] 1 1 1 2 3 1 2 1 1 1 ...
  .. attr(*, "label")= chr "Masked variance pseudo-PSU"
 $ sdmvstra      : num [1:5615] 91 94 92 96 90 99 92 95 92 97 ...
  .. attr(*, "label")= chr "Masked variance pseudo-stratum"
 $ indfmpir      : num [1:5615] 3.15 1.67 0.33 2.02 4.3 5 NA 0.05 5 0.87 ...
  .. attr(*, "label")= chr "Ratio of family income to poverty"
 $ BPXDI1       : num [1:5615] 82 56 80 70 NA NA 68 72 78 62 ...
  .. attr(*, "label")= chr "Diastolic: Blood pres (1st rdg) mm Hg"
 $ bmx bmi       : num [1:5615] 23.3 23.2 20.1 33.3 33.9 23.3 20.1 28.5 27.6 27.9 ...
  .. attr(*, "label")= chr "Body Mass Index (kg/m**2)"
 $ age          : num [1:5615] 22 44 21 43 80 34 51 80 55 35 ...
  .. attr(*, "label")= chr "Age at Interview in Years"
 $ marcat       : num [1:5615] 3 1 3 3 1 1 1 2 1 1 ...
  .. attr(*, "label")= chr "1=married 2=prev married 3=never married"
 $ bpxdi1_1     : num [1:5615] 82 56 80 70 0 0 68 72 78 62 ...
 $ agec         : num [1:5615] -24.36 -2.36 -25.36 -3.36 33.64 ...
 $ agecsq       : num [1:5615] 593.41 5.57 643.13 11.29 1131.65 ...
 $ descodes     : num [1:5615] 911 941 921 962 903 991 922 951 921 971 ...
 $ high_diastolic: num [1:5615] 0 0 0 0 NA NA 0 0 0 0 ...
> md.pattern(nhanesc12_sub)
      riagendr RIDRETH1 WTMEC2YR sdmvpsu sdmvstra age bpxdi1_1 agec agecsq descodes bmx bmi marcat
4430      1          1          1          1          1 1          1 1          1          1          1
371       1          1          1          1          1 1          1 1          1          1          1
372       1          1          1          1          1 1          1 1          1          1          1
60        1          1          1          1          1 1          1 1          1          1          1
231       1          1          1          1          1 1          1 1          1          1          0
32        1          1          1          1          1 1          1 1          1          1          0
21        1          1          1          1          1 1          1 1          1          1          0
8         1          1          1          1          1 1          1 1          1          1          0
49        1          1          1          1          1 1          1 1          1          0          1
12        1          1          1          1          1 1          1 1          1          0          1
17        1          1          1          1          1 1          1 1          1          0          1
4         1          1          1          1          1 1          1 1          1          0          1
7         1          1          1          1          1 1          1 1          1          0          0
1         1          1          1          1          1 1          1 1          1          0          0
      0          0          0          0          0 0          0 0          0          0          90          300
```

```

      BPXDI1 high_diastolic indfmpir
4430    1          1          1    0
371     1          1          0    1
372     0          0          1    2
60      0          0          0    3
231     1          1          1    1
32      1          1          0    2
21      0          0          1    3
8       0          0          0    4
49      1          1          1    1
12      1          1          0    2
17      0          0          1    3
4       0          0          0    4
7       1          1          1    2
1       0          0          1    4
      483          483          487 1843
>
> # Set survey design
> nhanessvy <- svydesign(strata=~sdmvstra, id=~sdmvpsu, weights=~WTMEC2YR, data=nhanesc12_sub, nest=T)

> # Complete Case Analysis
> # Obtain means for 3 continuous variables without imputation, Table 12.3
> print(ex12_3a <- svymean(~bmxbmi, nhanessvy, se=T, na.rm=T, ci=T))
      mean  SE
bmxbmi 28.6 0.21
> show(ex12_3b <- svymean(~BPXDI1, nhanessvy, se=T, na.rm=T, ci=T))
      mean  SE
BPXDI1 71.4 0.51
> show(ex12_3c <- svymean(~indfmpir, nhanessvy, se=T, na.rm=T, ci=T))
      mean  SE
indfmpir 2.86 0.11

> # High Blood Pressure Mean, Complete Case, Table 12.4
> (ex12_4 <- svymean(~factor(high_diastolic),nhanessvy, se=T, na.rm=T, deff=T, ci=T, keep.vars=T))
      mean      SE DEff
factor(high_diastolic)0 0.93937 0.00793 5.67
factor(high_diastolic)1 0.06063 0.00793 5.67
> confint(ex12_4)
      2.5 %  97.5 %
factor(high_diastolic)0 0.92382 0.95491
factor(high_diastolic)1 0.04509 0.07618

> # Logistic Regression Complete Case Analysis with Design Correction, Table 12.5
> mod12_5 <- svyglm(high_diastolic ~ factor(RIDRETH1) + factor(riagendr)+ agec + agecsq,family=quasibinomial,
design=nhanessvy)
> summary(mod12_5)

Call:
svyglm(formula = high_diastolic ~ factor(RIDRETH1) + factor(riagendr) +
      agec + agecsq, design = nhanessvy, family = quasibinomial)

Survey design:
svydesign(strata = ~sdmvstra, id = ~sdmvpsu, weights = ~WTMEC2YR,
      data = nhanesc12_sub, nest = T)

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-2.248873	0.198142	-11.35	4.9e-07	***
factor(RIDRETH1)2	-0.724205	0.245430	-2.95	0.01452	*
factor(RIDRETH1)3	0.131207	0.224460	0.58	0.57180	
factor(RIDRETH1)4	0.654740	0.246514	2.66	0.02406	*
factor(RIDRETH1)5	0.050763	0.244587	0.21	0.83975	
factor(riagendr)2	-0.549726	0.207123	-2.65	0.02414	*
agec	0.008290	0.006944	1.19	0.26009	
agecsq	-0.001624	0.000276	-5.89	0.00015	***

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

(Dispersion parameter for quasibinomial family taken to be 1.02)

Number of Fisher Scoring iterations: 6

>

```
> # Imputation Method 1 with Design Variables in MI Model, return to original data without the CC high diastolic blood pressure
```

```
>  
> nhanesc12_subm1 <- nhanesc12
```

```
> summary(nhanesc12_subm1)
```

riagendr	RIDRETH1	WTMEC2YR	sdmvpsu	sdmvstra	indfmpir
Min. :1.00	Min. :1.0	Min. : 4413	Min. :1.00	Min. : 90.0	Min. :0.0
1st Qu.:1.00	1st Qu.:3.0	1st Qu.: 16173	1st Qu.:1.00	1st Qu.: 92.0	1st Qu.:0.9
Median :2.00	Median :3.0	Median : 24567	Median :2.00	Median : 96.0	Median :1.8
Mean :1.51	Mean :3.3	Mean : 41318	Mean :1.64	Mean : 95.9	Mean :2.4
3rd Qu.:2.00	3rd Qu.:4.0	3rd Qu.: 45238	3rd Qu.:2.00	3rd Qu.: 99.0	3rd Qu.:3.9
Max. :2.00	Max. :5.0	Max. :222580	Max. :3.00	Max. :103.0	Max. :5.0
					NA's :487

BPXDI1	bmbxmi	age	marcat	bpxdi1_1	agec
Min. : 0.0	Min. :13.4	Min. :18.0	Min. :1.00	Min. : 0.0	Min. :-28.36
1st Qu.: 64.0	1st Qu.:23.8	1st Qu.:31.0	1st Qu.:1.00	1st Qu.: 60.0	1st Qu.: -15.36
Median : 72.0	Median :27.4	Median :47.0	Median :1.00	Median : 70.0	Median : 0.64
Mean : 70.7	Mean :28.6	Mean :47.2	Mean :1.65	Mean : 64.7	Mean : 0.81
3rd Qu.: 78.0	3rd Qu.:32.0	3rd Qu.:62.0	3rd Qu.:2.00	3rd Qu.: 78.0	3rd Qu.: 15.64
Max. :120.0	Max. :82.1	Max. :80.0	Max. :3.00	Max. :120.0	Max. : 33.64
NA's :483	NA's :90		NA's :300		

agecsq	descode	high_diastolic
Min. : 0.1	Min. : 901	Min. :0.0
1st Qu.: 58.4	1st Qu.: 922	1st Qu.:0.0
Median : 244.6	Median : 961	Median :0.0
Mean : 345.1	Mean : 960	Mean :0.1
3rd Qu.: 558.8	3rd Qu.: 992	3rd Qu.:0.0
Max. :1131.6	Max. :1032	Max. :1.0
		NA's :483

```
> names(nhanesc12_subm1)
```

[1] "riagendr"	"RIDRETH1"	"WTMEC2YR"	"sdmvpsu"	"sdmvstra"
[6] "indfmpir"	"BPXDI1"	"bmbxmi"	"age"	"marcat"
[11] "bpxdi1_1"	"agec"	"agecsq"	"descode"	"high_diastolic"

```
> # Subset variables by number position in data set , drop age and high_diastolic since will be redone once bpxdi1 is complete
```

```
> (nhanesc12_subset <- nhanesc12[, c(1,2,3,4,5,6,7,8,10,12,13,14)])
```

```
# A tibble: 5,615 × 12
```

	riagendr	RIDRETH1	WTMEC2YR	sdmvpsu	sdmvstra	indfmpir	BPXDI1	bmbxmi	marcat	agec	agecsq	descode
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	3	104237.	1	91	3.15	82	23.3	3	-24.4	593.	911
2	2	3	127965.	1	94	1.67	56	23.2	1	-2.36	5.57	941
3	1	5	14784.	1	92	0.33	80	20.1	3	-25.4	643.	921
4	2	4	27123.	2	96	2.02	70	33.3	3	-3.36	11.3	962
5	1	3	27336.	3	90	4.3	NA	33.9	1	33.6	1132.	903
6	2	3	54203.	1	99	5	NA	23.3	1	-12.4	153.	991
7	1	5	7851.	2	92	NA	68	20.1	1	4.64	21.5	922
8	1	3	13190.	1	95	0.05	72	28.5	2	33.6	1132.	951
9	1	5	17115.	1	92	5	78	27.6	1	8.64	74.6	921
10	1	3	22616.	1	97	0.87	62	27.9	1	-11.4	129.	971

```
# i 5,605 more rows
```

```
# i Use `print(n = ...)` to see more rows
```

```
> summary(nhanesc12_subset)
```

riagendr	RIDRETH1	WTMEC2YR	sdmvpsu	sdmvstra	indfmpir
Min. :1.00	Min. :1.0	Min. : 4413	Min. :1.00	Min. : 90.0	Min. :0.0
1st Qu.:1.00	1st Qu.:3.0	1st Qu.: 16173	1st Qu.:1.00	1st Qu.: 92.0	1st Qu.:0.9
Median :2.00	Median :3.0	Median : 24567	Median :2.00	Median : 96.0	Median :1.8
Mean :1.51	Mean :3.3	Mean : 41318	Mean :1.64	Mean : 95.9	Mean :2.4
3rd Qu.:2.00	3rd Qu.:4.0	3rd Qu.: 45238	3rd Qu.:2.00	3rd Qu.: 99.0	3rd Qu.:3.9

```

Max. :2.00 Max. :5.0 Max. :222580 Max. :3.00 Max. :103.0 Max. :5.0
NA's :487
BPXDI1 bmx bmi marcat agec agecsq descodes
Min. : 0.0 Min. :13.4 Min. :1.00 Min. :-28.36 Min. : 0.1 Min. : 901
1st Qu.: 64.0 1st Qu.:23.8 1st Qu.:1.00 1st Qu.: -15.36 1st Qu.: 58.4 1st Qu.: 922
Median : 72.0 Median :27.4 Median :1.00 Median : 0.64 Median : 244.6 Median : 961
Mean : 70.7 Mean :28.6 Mean :1.65 Mean : 0.81 Mean : 345.1 Mean : 960
3rd Qu.: 78.0 3rd Qu.:32.0 3rd Qu.:2.00 3rd Qu.: 15.64 3rd Qu.: 558.8 3rd Qu.: 992
Max. :120.0 Max. :82.1 Max. :3.00 Max. : 33.64 Max. :1131.6 Max. :1032
NA's :483 NA's :90 NA's :300

```

```

>
> # Use mice to impute and specify type of default method for imputation models
> # Run without custom predictor matrix first, see which variables might not be used in final model

```

```

> impm1 <- mice(nhanesc12_subset, n.imp=5, seed=2024, defaultMethod=c("norm", "logreg", "polyreg"))

```

```

iter imp variable
 1  1  indfmpir BPXDI1 bmx bmi marcat
 1  2  indfmpir BPXDI1 bmx bmi marcat
 1  3  indfmpir BPXDI1 bmx bmi marcat
 1  4  indfmpir BPXDI1 bmx bmi marcat
 1  5  indfmpir BPXDI1 bmx bmi marcat
 2  1  indfmpir BPXDI1 bmx bmi marcat
 2  2  indfmpir BPXDI1 bmx bmi marcat
 2  3  indfmpir BPXDI1 bmx bmi marcat
 2  4  indfmpir BPXDI1 bmx bmi marcat
 2  5  indfmpir BPXDI1 bmx bmi marcat
 3  1  indfmpir BPXDI1 bmx bmi marcat
 3  2  indfmpir BPXDI1 bmx bmi marcat
 3  3  indfmpir BPXDI1 bmx bmi marcat
 3  4  indfmpir BPXDI1 bmx bmi marcat
 3  5  indfmpir BPXDI1 bmx bmi marcat
 4  1  indfmpir BPXDI1 bmx bmi marcat
 4  2  indfmpir BPXDI1 bmx bmi marcat
 4  3  indfmpir BPXDI1 bmx bmi marcat
 4  4  indfmpir BPXDI1 bmx bmi marcat
 4  5  indfmpir BPXDI1 bmx bmi marcat
 5  1  indfmpir BPXDI1 bmx bmi marcat
 5  2  indfmpir BPXDI1 bmx bmi marcat
 5  3  indfmpir BPXDI1 bmx bmi marcat
 5  4  indfmpir BPXDI1 bmx bmi marcat
 5  5  indfmpir BPXDI1 bmx bmi marcat

```

Warning message:

Number of logged events: 1

```

> impm1$predictorMatrix

```

```

riagendr RIDRETH1 WTMEC2YR sdmvpsu sdmvstra indfmpir BPXDI1 bmx bmi marcat agec agecsq
riagendr      0      1      1      1      1      1      1      1      1      1      1
RIDRETH1      1      0      1      1      1      1      1      1      1      1      1
WTMEC2YR      1      1      0      1      1      1      1      1      1      1      1
sdmvpsu       1      1      1      0      1      1      1      1      1      1      1
sdmvstra      1      1      1      1      0      1      1      1      1      1      1
indfmpir      1      1      1      1      1      0      1      1      1      1      1
BPXDI1        1      1      1      1      1      1      0      1      1      1      1
bmx bmi       1      1      1      1      1      1      1      0      1      1      1
marcat        1      1      1      1      1      1      1      1      0      1      1
agec          1      1      1      1      1      1      1      1      1      0      1
agecsq        1      1      1      1      1      1      1      1      1      1      0

```

```

descodes      1      1      1      1      1      1      1      1      1      1      1      1
descodes
riagendr      0
RIDRETH1      0
WTMEC2YR      0
sdmvpsu       0
sdmvstra      0
indfmpir      0
BPXDI1        0
bmxbmi        0
marcat        0
agec          0
agecsq        0
descodes      0
> summary(imp1)
Class: mids
Number of multiple imputations: 5
Imputation methods:
riagendr RIDRETH1 WTMEC2YR  sdmvpsu sdmvstra indfmpir  BPXDI1  bmxbmi  marcat  agec  agecsq
  ""      ""      ""      ""      ""      "norm"  "norm"  "norm"  "norm"  ""   ""

descodes
  ""

PredictorMatrix:
      riagendr RIDRETH1 WTMEC2YR sdmvpsu sdmvstra indfmpir BPXDI1 bmxbmi marcat agec agecsq
riagendr      0      1      1      1      1      1      1      1      1      1      1
RIDRETH1      1      0      1      1      1      1      1      1      1      1      1
WTMEC2YR      1      1      0      1      1      1      1      1      1      1      1
sdmvpsu       1      1      1      0      1      1      1      1      1      1      1
sdmvstra      1      1      1      1      0      1      1      1      1      1      1
indfmpir      1      1      1      1      1      0      1      1      1      1      1

descodes
riagendr      0
RIDRETH1      0
WTMEC2YR      0
sdmvpsu       0
sdmvstra      0
indfmpir      0
Number of logged events: 1
  it im dep      meth      out
1 0 0      collinear descodes
>
> # Add a predictor matrix to control imputation model predictors for each imputed variable
> pred <- imp1$predictorMatrix
> pred[,"sdmvpsu"] <- 0
> pred[,"sdmvstra"] <- 0
> pred[,"descodes"] <- 1
> pred[,"WTMEC2YR"] <- 1
>
> pred #note that certain variables are set to 0 for this matrix
      riagendr RIDRETH1 WTMEC2YR sdmvpsu sdmvstra indfmpir BPXDI1 bmxbmi marcat agec agecsq
riagendr      0      1      1      0      0      1      1      1      1      1      1
RIDRETH1      1      0      1      0      0      1      1      1      1      1      1
WTMEC2YR      1      1      1      0      0      1      1      1      1      1      1
sdmvpsu       1      1      1      0      0      1      1      1      1      1      1
sdmvstra      1      1      1      0      0      1      1      1      1      1      1
indfmpir      1      1      1      0      0      0      1      1      1      1      1
BPXDI1        1      1      1      0      0      1      0      1      1      1      1
bmxbmi        1      1      1      0      0      1      1      0      1      1      1
marcat        1      1      1      0      0      1      1      1      0      1      1

```

```

agec      1      1      1      0      0      1      1      1      1      0      1
agecsq    1      1      1      0      0      1      1      1      1      1      0
descodes  1      1      1      0      0      1      1      1      1      1      1

```

```
descodes
```

```

riagendr  1
RIDRETH1  1
WTMEC2YR  1
sdmvpsu   1
sdmvstra  1
indfmpir  1
BPXDI1    1
bmxbmi    1
marcat     1
agec      1
agecsq    1
descodes  1

```

```

>
> impm1_pred <- mice(nhanesc12_subset, pred=pred, n.imp=5, seed=2016,
defaultMethod=c("norm","logreg","polyreg"))

```

```

iter imp variable
  1  1  indfmpir  BPXDI1  bmxbmi  marcat
  1  2  indfmpir  BPXDI1  bmxbmi  marcat
  1  3  indfmpir  BPXDI1  bmxbmi  marcat
  1  4  indfmpir  BPXDI1  bmxbmi  marcat
  1  5  indfmpir  BPXDI1  bmxbmi  marcat
  2  1  indfmpir  BPXDI1  bmxbmi  marcat
  2  2  indfmpir  BPXDI1  bmxbmi  marcat
  2  3  indfmpir  BPXDI1  bmxbmi  marcat
  2  4  indfmpir  BPXDI1  bmxbmi  marcat
  2  5  indfmpir  BPXDI1  bmxbmi  marcat
  3  1  indfmpir  BPXDI1  bmxbmi  marcat
  3  2  indfmpir  BPXDI1  bmxbmi  marcat
  3  3  indfmpir  BPXDI1  bmxbmi  marcat
  3  4  indfmpir  BPXDI1  bmxbmi  marcat
  3  5  indfmpir  BPXDI1  bmxbmi  marcat
  4  1  indfmpir  BPXDI1  bmxbmi  marcat
  4  2  indfmpir  BPXDI1  bmxbmi  marcat
  4  3  indfmpir  BPXDI1  bmxbmi  marcat
  4  4  indfmpir  BPXDI1  bmxbmi  marcat
  4  5  indfmpir  BPXDI1  bmxbmi  marcat
  5  1  indfmpir  BPXDI1  bmxbmi  marcat
  5  2  indfmpir  BPXDI1  bmxbmi  marcat
  5  3  indfmpir  BPXDI1  bmxbmi  marcat
  5  4  indfmpir  BPXDI1  bmxbmi  marcat
  5  5  indfmpir  BPXDI1  bmxbmi  marcat

```

```
> summary(impm1_pred)
```

```
Class: mids
```

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

```
Imputation methods:
```

```

riagendr  RIDRETH1  WTMEC2YR  sdmvpsu  sdmvstra  indfmpir  BPXDI1  bmxbmi  marcat  agec  agecsq
  ""      ""      ""      ""      ""      "norm"  "norm"  "norm"  "norm"  ""   ""
descodes
  ""

```

```
PredictorMatrix:
```

```

riagendr  RIDRETH1  WTMEC2YR  sdmvpsu  sdmvstra  indfmpir  BPXDI1  bmxbmi  marcat  agec  agecsq
riagendr  0      1      1      0      0      1      1      1      1      1      1
RIDRETH1  1      0      1      0      0      1      1      1      1      1      1
WTMEC2YR  1      1      0      0      0      1      1      1      1      1      1

```

```

sdmvpsu      1      1      1      0      0      1      1      1      1      1      1
sdmvstra     1      1      1      0      0      1      1      1      1      1      1
indfmpir     1      1      1      0      0      0      1      1      1      1      1
  descodes
riagendr     1
RIDRETH1     1
WTMEC2YR     1
sdmvpsu      1
sdmvstra     1
indfmpir     1

```

```
> # Note small differences between 1st and 2nd implicate
```

```
> summary(mice::complete( impm1_pred , 1 )$BPXDI1)
```

```

Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0   64.0   72.0   70.7   78.0  120.0

```

```
> summary(mice::complete( impm1_pred , 2 )$BPXDI1)
```

```

Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0   64.0   72.0   70.8   78.0  120.0

```

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

```
> list_of_data_frames <- list(
```

```
+ mice::complete(impm1_pred , 1 ) ,
```

```
+ mice::complete( impm1_pred , 2 ) ,
```

```
+ mice::complete( impm1_pred , 3 ) ,
```

```
+ mice::complete( impm1_pred , 4 ) ,
```

```
+ mice::complete( impm1_pred , 5 ))
```

```
> # Multiply-imputed design statement
```

```
> desnhanesm1 <- svydesign(id = ~sdmvpsu, strat = ~sdmvstra , weight = ~WTMEC2YR ,
```

```
+ data = imputationList( list_of_data_frames ) , nest = TRUE )
```

```
> desnhanesm1
```

```
Multiple (5) imputations: svydesign(id = ~sdmvpsu, strat = ~sdmvstra, weight = ~WTMEC2YR,
  data = imputationList(list_of_data_frames), nest = TRUE)
```

```
> # Add a new variable to desnhanesm1 using imputed bpxdi1 values
```

```
> desnhanesm1 <- update(desnhanesm1, high_diastolic = ifelse(BPXDI1 >=90,1,0) )
```

```
> (desnhanesm1)
```

```
Multiple (5) imputations: svydesign(id = ~sdmvpsu, strat = ~sdmvstra, weight = ~WTMEC2YR,
  data = imputationList(list_of_data_frames), nest = TRUE)
```

```
> # Combined mean high blood pressure with design adjustment, Table 12.4
```

```
> fitm1_mean <- MIcombine(with(desnhanesm1, svymean(~factor(high_diastolic), se=T, ci=T, na.rm=T)))
```

```
> fitm1_mean
```

```
Multiple imputation results:
```

```
  with(desnhanesm1, svymean(~factor(high_diastolic), se = T, ci = T,
na.rm = T))
```

```
  MIcombine.default(with(desnhanesm1, svymean(~factor(high_diastolic),
se = T, ci = T, na.rm = T)))
```

```
          results          se
```

```
factor(high_diastolic)0 0.93795 0.00773
```

```
factor(high_diastolic)1 0.06205 0.00773
```

```
> confint(fitm1_mean)
```

```
          2.5 % 97.5 %
```

```
factor(high_diastolic)0 0.9228 0.9531
```

```
factor(high_diastolic)1 0.0469 0.0772
```

```

> # Run design based logistic model with svyglm using 5 imputed data sets contained in desnhanesm1, Table 12.5
> fitm1 <- MIcombine(with(desnhanesm1, svyglm(factor(high_diastolic) ~ factor(RIDRETH1) + factor(riagendr) +
agec + agecsq, family=quasibinomial)))
> fitm1
Multiple imputation results:
  with(desnhanesm1, svyglm(factor(high_diastolic) ~ factor(RIDRETH1) +
factor(riagendr) + agec + agecsq, family = quasibinomial))
  MIcombine.default(with(desnhanesm1, svyglm(factor(high_diastolic) ~
factor(RIDRETH1) + factor(riagendr) + agec + agecsq, family = quasibinomial)))
              results      se
(Intercept)  -2.242033 0.2131718
factor(RIDRETH1)2 -0.553521 0.3382193
factor(RIDRETH1)3  0.149216 0.2418729
factor(RIDRETH1)4  0.612931 0.2544578
factor(RIDRETH1)5  0.106958 0.2742154
factor(riagendr)2 -0.492720 0.1897702
agec          0.008941 0.0067064
agecsq       -0.001697 0.0002728

> # Obtain means for 3 continuous variables imputed using method 1, Table 12.3 (by each replicate)
> fitm1_ex12_3 <- with(desnhanesm1, svymean(~bmxbmi + BPXDI1 + indfmpir), se=T, na.rm=T, ci=T )
> fitm1_ex12_3
[[1]]
      mean  SE
bmxbmi 28.62 0.21
BPXDI1 71.30 0.46
indfmpir 2.86 0.10

[[2]]
      mean  SE
bmxbmi 28.63 0.21
BPXDI1 71.45 0.47
indfmpir 2.84 0.10

[[3]]
      mean  SE
bmxbmi 28.61 0.21
BPXDI1 71.36 0.45
indfmpir 2.85 0.10

[[4]]
      mean  SE
bmxbmi 28.63 0.21
BPXDI1 71.36 0.47
indfmpir 2.85 0.10

[[5]]
      mean  SE
bmxbmi 28.63 0.21
BPXDI1 71.32 0.44
indfmpir 2.84 0.10

attr(,"call")
with(desnhanesm1, svymean(~bmxbmi + BPXDI1 + indfmpir), se = T,
      na.rm = T, ci = T)
>

```

```

> # Imputation 2 NO design variables in models
> # Return to original data without the CC high diastolic blood pressure
>
> nhanesc12_submeth2 <- nhanesc12
> names(nhanesc12_submeth2)
 [1] "riagendr"      "RIDRETH1"      "WTMEC2YR"      "sdmvpsu"      "sdmvstra"
 [6] "indfmpir"     "BPXDI1"        "bmxbmi"        "age"          "marcat"
[11] "bpxdi1_1"     "agec"          "agecsq"        "decode"       "high_diastolic"
>
> # Subset variables by number position in data set , drop high_diastolic for imputations
> (nhanesc12_subset <- nhanesc12[, c(1,2,3,4,5,6,7,8,10, 12,13,14)])
# A tibble: 5,615 × 12
  riagendr RIDRETH1 WTMEC2YR sdmvpsu sdmvstra indfmpir BPXDI1 bmxbmi marcat agec agecsq decode
  <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1         1         3 104237.         1         91         3.15         82         23.3         3 -24.4         593.         911
2         2         3 127965.         1         94         1.67         56         23.2         1  -2.36          5.57         941
3         1         5 14784.          1         92         0.33         80         20.1         3 -25.4         643.         921
4         2         4 27123.          2         96         2.02         70         33.3         3  -3.36          11.3         962
5         1         3 27336.          3         90         4.3          NA         33.9         1  33.6         1132.         903
6         2         3 54203.          1         99         5           NA         23.3         1 -12.4          153.         991
7         1         5 7851.           2         92         NA          68         20.1         1   4.64          21.5         922
8         1         3 13190.          1         95         0.05         72         28.5         2  33.6         1132.         951
9         1         5 17115.          1         92         5           78         27.6         1   8.64          74.6         921
10        1         3 22616.          1         97         0.87         62         27.9         1 -11.4          129.         971
# i 5,605 more rows
# i Use `print(n = ...)` to see more rows
> summary(nhanesc12_subset)
  riagendr      RIDRETH1      WTMEC2YR      sdmvpsu      sdmvstra      indfmpir
Min.   :1.00   Min.    :1.0   Min.    : 4413   Min.    :1.00   Min.    : 90.0   Min.    :0.0
1st Qu.:1.00   1st Qu.:3.0   1st Qu.: 16173   1st Qu.:1.00   1st Qu.: 92.0   1st Qu.:0.9
Median :2.00   Median :3.0   Median : 24567   Median :2.00   Median : 96.0   Median :1.8
Mean   :1.51   Mean    :3.3   Mean    : 41318   Mean    :1.64   Mean    : 95.9   Mean    :2.4
3rd Qu.:2.00   3rd Qu.:4.0   3rd Qu.: 45238   3rd Qu.:2.00   3rd Qu.: 99.0   3rd Qu.:3.9
Max.   :2.00   Max.    :5.0   Max.    :222580   Max.    :3.00   Max.    :103.0   Max.    :5.0
                                     NA's    :487

  BPXDI1      bmxbmi      marcat      agec      agecsq      decode
Min.    : 0.0   Min.    :13.4   Min.    :1.00   Min.    :-28.36   Min.    : 0.1   Min.    : 901
1st Qu.: 64.0   1st Qu.:23.8   1st Qu.:1.00   1st Qu.: -15.36   1st Qu.: 58.4   1st Qu.: 922
Median : 72.0   Median :27.4   Median :1.00   Median : 0.64   Median : 244.6   Median : 961
Mean    : 70.7   Mean    :28.6   Mean    :1.65   Mean    : 0.81   Mean    : 345.1   Mean    : 960
3rd Qu.: 78.0   3rd Qu.:32.0   3rd Qu.:2.00   3rd Qu.: 15.64   3rd Qu.: 558.8   3rd Qu.: 992
Max.    :120.0   Max.    :82.1   Max.    :3.00   Max.    : 33.64   Max.    :1131.6   Max.    :1032
NA's    :483    NA's    :90    NA's    :300
>
> nhanesc12_subm2 <- nhanesc12_subset
>
> # Add a predictor matrix to control imputation model predictors for each imputed variable
> pred <- impm2$predictorMatrix
Error: object 'impm2' not found
> pred[,"sdmvpsu"] <- 0
> pred[,"sdmvstra"] <- 0
> pred[,"decode"] <- 0
> pred[,"WTMEC2YR"] <- 0
> impm2$predictorMatrix
Error: object 'impm2' not found
>
> pred
  riagendr RIDRETH1 WTMEC2YR sdmvpsu sdmvstra indfmpir BPXDI1 bmxbmi marcat agec agecsq
riagendr      0         1         0         0         0         1         1         1         1         1         1

```

```

RIDRETH1      1      0      0      0      0      1      1      1      1      1      1
WTMEC2YR      1      1      0      0      0      1      1      1      1      1      1
sdmvpsu       1      1      0      0      0      1      1      1      1      1      1
sdmvstra      1      1      0      0      0      1      1      1      1      1      1
indfmpir      1      1      0      0      0      0      1      1      1      1      1
BPXDI1        1      1      0      0      0      1      0      1      1      1      1
bmxbmi        1      1      0      0      0      1      1      0      1      1      1
marcat        1      1      0      0      0      1      1      1      0      1      1
agec          1      1      0      0      0      1      1      1      1      0      1
agecsq        1      1      0      0      0      1      1      1      1      1      0
descodes      1      1      0      0      0      1      1      1      1      1      1

```

```
descodes
```

```

riagendr      0
RIDRETH1      0
WTMEC2YR      0
sdmvpsu       0
sdmvstra      0
indfmpir      0
BPXDI1        0
bmxbmi        0
marcat        0
agec          0
agecsq        0
descodes      0

```

```

>
> impm2_pred <- mice(nhanesc12_subm2, pred=pred, n.imp=5, seed=2024,
defaultMethod=c("norm", "logreg", "polyreg"))

```

```
iter imp variable
```

```

 1  1  indfmpir  BPXDI1  bmxbmi  marcat
 1  2  indfmpir  BPXDI1  bmxbmi  marcat
 1  3  indfmpir  BPXDI1  bmxbmi  marcat
 1  4  indfmpir  BPXDI1  bmxbmi  marcat
 1  5  indfmpir  BPXDI1  bmxbmi  marcat
 2  1  indfmpir  BPXDI1  bmxbmi  marcat
 2  2  indfmpir  BPXDI1  bmxbmi  marcat
 2  3  indfmpir  BPXDI1  bmxbmi  marcat
 2  4  indfmpir  BPXDI1  bmxbmi  marcat
 2  5  indfmpir  BPXDI1  bmxbmi  marcat
 3  1  indfmpir  BPXDI1  bmxbmi  marcat
 3  2  indfmpir  BPXDI1  bmxbmi  marcat
 3  3  indfmpir  BPXDI1  bmxbmi  marcat
 3  4  indfmpir  BPXDI1  bmxbmi  marcat
 3  5  indfmpir  BPXDI1  bmxbmi  marcat
 4  1  indfmpir  BPXDI1  bmxbmi  marcat
 4  2  indfmpir  BPXDI1  bmxbmi  marcat
 4  3  indfmpir  BPXDI1  bmxbmi  marcat
 4  4  indfmpir  BPXDI1  bmxbmi  marcat
 4  5  indfmpir  BPXDI1  bmxbmi  marcat
 5  1  indfmpir  BPXDI1  bmxbmi  marcat
 5  2  indfmpir  BPXDI1  bmxbmi  marcat
 5  3  indfmpir  BPXDI1  bmxbmi  marcat
 5  4  indfmpir  BPXDI1  bmxbmi  marcat
 5  5  indfmpir  BPXDI1  bmxbmi  marcat

```

```
> summary(impm2_pred)
```

```
Class: mids
```

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

```
Imputation methods:
```

```
riagendr RIDRETH1 WTMEC2YR sdmvpsu sdmvstra indfmpir BPXDI1 bmxbmi marcat agec agecsq
```

```

    ""      ""      ""      ""      ""      "norm"  "norm"  "norm"  "norm"      ""      ""
descode
    ""
PredictorMatrix:
      riagendr RIDRETH1 WTMEC2YR sdmvpsu sdmvstra indfmpir BPXDI1 bmxbmi marcat agec agecsq
riagendr      0         1         0         0         0         1         1         1         1         1         1
RIDRETH1      1         0         0         0         0         1         1         1         1         1         1
WTMEC2YR      1         1         0         0         0         1         1         1         1         1         1
sdmvpsu       1         1         0         0         0         1         1         1         1         1         1
sdmvstra      1         1         0         0         0         1         1         1         1         1         1
indfmpir      1         1         0         0         0         0         1         1         1         1         1
descode
riagendr      0
RIDRETH1      0
WTMEC2YR      0
sdmvpsu       0
sdmvstra      0
indfmpir      0
>
> # Small differences between 1st and 2nd implicate
> summary(mice::complete( impm2_pred , 1 )$BPXDI1)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.0   64.0   72.0   70.8   78.0  120.0
> summary(mice::complete( impm2_pred , 2 )$BPXDI1)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.0   64.0   71.7   70.7   78.0  120.0

> # Combine all five data.frame objects into a list
> list_of_data_frames2 <-list(
+ mice::complete( impm2_pred , 1 ) ,
+ mice::complete( impm2_pred , 2 ) ,
+ mice::complete( impm2_pred , 3 ) ,
+ mice::complete( impm2_pred , 4 ) ,
+ mice::complete( impm2_pred , 5 )
+ )

> # Multiply-imputed design statement
> desnhanesm2 <-svydesign(id = ~ sdmvpsu, strat = ~ sdmvstra , weight = ~ WTMEC2YR ,
+ data = imputationList( list_of_data_frames2 ), nest = TRUE)
> desnhanesm2
Multiple (5) imputations: svydesign(id = ~sdmvpsu, strat = ~sdmvstra, weight = ~WTMEC2YR,
  data = imputationList(list_of_data_frames2), nest = TRUE)

> # Add a new variable to desnhanesm2 using imputed blood pressure values
> desnhanesm2 <- update( desnhanesm2, high_diastolic = ifelse(BPXDI1 >=90,1,0))
> (desnhanesm2)
Multiple (5) imputations: svydesign(id = ~sdmvpsu, strat = ~sdmvstra, weight = ~WTMEC2YR,
  data = imputationList(list_of_data_frames2), nest = TRUE)

```

```

> # Combined mean high blood pressure with NO design adjustment, Table 12.4
> fitm2_mean <- MIcombine(with(desnhanesm2, svymean(~factor(high_diastolic), se=T, na.rm=T, ci=T )))

> fitm2_mean
Multiple imputation results:
  with(desnhanesm2, svymean(~factor(high_diastolic), se = T, na.rm = T,
ci = T))
  MIcombine.default(with(desnhanesm2, svymean(~factor(high_diastolic),
se = T, na.rm = T, ci = T)))
                results      se
factor(high_diastolic)0 0.93931 0.006851
factor(high_diastolic)1 0.06069 0.006851

> # Run design based logistic model with svyglm using 5 imputed data sets contained in desnhanesm2, Table 12.5
> fitm2 <- MIcombine(with(desnhanesm2, svyglm (factor(high_diastolic) ~ factor(RIDRETH1) + factor(riagendr) +
agec + agecsq, family=quasibinomial)))

> (fitm2)
Multiple imputation results:
  with(desnhanesm2, svyglm(factor(high_diastolic) ~ factor(RIDRETH1) +
factor(riagendr) + agec + agecsq, family = quasibinomial))
  MIcombine.default(with(desnhanesm2, svyglm(factor(high_diastolic) ~
factor(RIDRETH1) + factor(riagendr) + agec + agecsq, family = quasibinomial)))
                results      se
(Intercept)      -2.246027 0.2209162
factor(RIDRETH1)2 -0.627261 0.2862377
factor(RIDRETH1)3  0.149691 0.2387049
factor(RIDRETH1)4  0.644521 0.2527741
factor(RIDRETH1)5  0.064117 0.2469976
factor(riagendr)2 -0.541073 0.2045024
agec              0.007834 0.0066280
agecsq           -0.001691 0.0003141

> # Obtain means for 3 continuous variables imputed using method 2 (NO DESIGN) , Table 12.3 (by each replicate)
> fitm2_ex12_3 <- with(desnhanesm2, svymean(~bmx bmi + BPXDI1 + indfmpir, se=T, na.rm=T, ci=T ))
> fitm2_ex12_3
[[1]]
      mean  SE
bmx bmi  28.61 0.21
BPXDI1   71.41 0.43
indfmpir  2.83 0.10

[[2]]
      mean  SE
bmx bmi  28.62 0.21
BPXDI1   71.25 0.42
indfmpir  2.83 0.10

[[3]]
      mean  SE
bmx bmi  28.62 0.21
BPXDI1   71.31 0.46
indfmpir  2.83 0.10

[[4]]
      mean  SE
bmx bmi  28.62 0.21

```

```
BPXDI1 71.36 0.44
indfmpir 2.82 0.10
```

```
[[5]]
```

```
      mean  SE
bmx bmi 28.62 0.21
BPXDI1 71.27 0.47
indfmpir 2.82 0.10
```

```
attr(,"call")
```

```
with(desnhanesm2, svymean(~bmx bmi + BPXDI1 + indfmpir, se = T,
  na.rm = T, ci = T))
```

```
> # Note: FEFI method not available in R Survey Package, see examples using other software such as SAS PROC SURVEYIMPUTE
```