ASDA3 Analysis Examples Replication Chapter 7

R Code

```
# Load packages and prepare data for analysis
library (survey)
library(haven)

setwd("P:/ASDA3/Replication R/Chapter 7")

load("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/nhanes1112.rdata")
nhanesdata <- nhanes1112

# Set key variables to values (0=NA for blood pressure, center age and create agecs2)
nhanesdata$bpxdi1_1 <-nhanesdata$bpxdi1
nhanesdata$bpxdi1_1 [nhanesdata$bpxdi1_1 ==0] <- NA
nhanesdata$agec <- (nhanesdata$age-46.36)
nhanesdata$agecsq <- (nhanesdata$agec*nhanesdata$agec)
nhanesdata$genderc <- factor(nhanesdata$riagendr , levels = 1: 2, labels =c("M", "F"))
summary(nhanesdata)

# Create factor variables for models
nhanesdata$racec <- factor(nhanesdata$ridreth1, levels = 1: 5 , labels =c("Mexican", "Other Hispanic", "White",
"Black", "Other"))
nhanesdata$marcatc <- factor(nhanesdata$marcat, levels = 1: 3, labels =c("Married", "Previously Married", "Never
Married"))
nhanesdata$edcatc <- factor(nhanesdata$edcat, levels = 1: 4, labels =c("0-11", "12", "13-15","16+"))
nhanesdata$bp_catc <- factor(nhanesdata$bp_cat, levels = 1: 4, labels =c("Normal", "Pre-HBP", "Stage 1
HBP","Stage 2 HBP"))

# Subpop indicator
nhanesdata$age18p <- ifelse(nhanesdata$age >=18,1,0)

# Survey design
nhanessvy2 <- svydesign(strata=~sdmvstra, id=~sdmvpsu, weights=~wtmec2yr, data=nhanesdata, nest=T)
subnhanes <- subset(nhanessvy2 , age >= 18)



# Table 7.1 Numbers from Bivariate Tests of Race, Marital Status, Gender, Centered Age
# Race/ethnicity
summary(ex75_race <- svyglm(bpxdi1_1 ~racec, design=subnhanes))
regTermTest(ex75_race, ~racec)

# Marital Status
summary(ex75_marital <- svyglm(bpxdi1_1 ~marcatc, design=subnhanes))
regTermTest(ex75_marital, ~marcatc)

# Gender
summary(ex75_sex <- svyglm(bpxdi1_1 ~genderc, design=subnhanes))
regTermTest(ex75_sex, ~genderc)

# Centered age
summary(ex75_age <- svyglm(bpxdi1_1 ~agec, design=subnhanes))

# 7.5.2 naive analysis, Table 7.2
(ex75_nowt <- lm(bpxdi1_1 ~ racec + marcatc + genderc + agec, data=nhanesdata, age >=18 ))
summary(ex75_nowt)
```

```
# 7.5.3 weighted linear regression without design variables, Table 7.3
(ex75_wt <- lm(bpxdi1_1 ~ racec + marcatc + genderc + agec, data= nhanesdata, age >=18, weight=wtmec2yr ))
summary(ex75_wt)

#  7.5.4 incorporating all design features, Table 7.4
summary(ex75_svyglm <- svyglm(bpxdi1_1 ~ racec + marcatc + genderc + agec, deff=T, design=subnhanes))
deff(ex75_svyglm) # note that these differ from Stata, see documentation for details
plot(ex75_svyglm) # this produces many plots, sift through for plots of interest in plot window

# Test if weights are informative, use new design structure with strata and psu but no weights
# Run model with same outcome but use weight and weight * each predictor
# Use regtermtest for joint test of weight and weight interactions

nhanessvy_nowt <- svydesign(strata=~sdmvstra, id=~sdmvpsu, data=nhanesdata, nest=T)
subnhanes_nowt <- subset(nhanessvy_nowt , age >= 18)

# 7.5.4  Run model and do test of weight and weight interactions
summary(ex75_nowt_svyglm <- svyglm(bpxdi1_1 ~ wtmec2yr + racec + genderc + agec +
wtmec2yr:(racec) + wtmec2yr:(genderc) + wtmec2yr:(marcatc) + wtmec2yr:agec , deff=T, design=subnhanes_nowt))
coef(ex75_nowt_svyglm)
summary(ex75_nowt_svyglm)

# Note: df differs from Stata but the set of variables p value still shows significant as a group, weights are
informative
regTermTest(ex75_nowt_svyglm, ~wtmec2yr + wtmec2yr:agec + wtmec2yr:racec + wtmec2yr:genderc + wtmec2yr:marcatc ,
df=NULL)

# Example 7.5 re-run initial model with informative weights and assess basic regression diagnostics
summary(ex75ini_svyglm <- svyglm(bpxdi1_1 ~ racec + marcatc + genderc + agec, deff=T, design=subnhanes))
deff(ex75ini_svyglm)

plot(ex75ini_svyglm) # basic plots, see documentation for various other plotting options available in R from svy
commands

# Example 7.5 with agecsq (agec squared), Table 7.5
summary(ex75_svyglm_agesq <- svyglm(bpxdi1_1 ~ racec + marcatc + genderc + agec + agecsq, deff=T,
design=subnhanes))
summary(ex75_svyglm_agesq)

plot(ex75_svyglm_agesq) # basic plots, see documentation for various other plotting options available in R from
svy commands

# Preliminary final model with race X age interactions
ex75_raceint <- svyglm(bpxdi1_1 ~ marcatc + genderc + factor(racec)*agec + factor(racec)*agecsq , subnhanes)
summary(ex75_raceint, df.resid=Inf)

# Joint test of race X age interactions
regTermTest(ex75_raceint, ~ factor(racec):agec + factor(racec):agecsq, df=Inf)

# Preliminary final model plus age vars X gender interactions
ex75_genint <- svyglm(bpxdi1_1 ~ marcatc + racec + factor(genderc)*agec + factor(genderc)*agecsq , subnhanes)
summary(ex75_genint, df.resid=Inf)

# Joint test of gender X age interactions
regTermTest(ex75_genint, ~ factor(genderc):agec + factor(genderc):agecsq, df=Inf, null=NULL)


# Final model including interactions of race X age, gender X age
ex75_final <- svyglm(bpxdi1_1 ~ marcatc + agec*factor(racec) + agecsq*factor(racec) + factor(genderc)*agec +
```

```
 factor(genderc)*agecsq, subnhanes)
summary(ex75_final, df.resid=Inf)


# Forthcoming!!! Check the website for examples of use of R packages for margins and plot of margins

# Model diagnostics
# Load svydiags package
library(svydiags)

# Note the variables names are slightly different than models previously run but are equivalent:
bpxdi1.1=bpxdi1_1 etc.
load("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/nhanes1112.rdata")

 nhanes1112$bpxdi1.1 <- nhanes1112$bpxdi1
 nhanes1112$bpxdi1.1[nhanes1112$bpxdi1 == 0] <- NA
 nhanes1112$agec <- nhanes1112$age - 46.36
 nhanes1112$agec2 <- nhanes1112$agec ^ 2

# Select needed variables and obtain complete case frame
nhanes.red <- nhanes1112[, c("bpxdi1.1", "ridreth1", "riagendr", "marcat", "agec", "agec2", "age18p",
"wtmec2yr", "sdmvstra", "sdmvpsu")]
nhanes.red <- nhanes.red[complete.cases(nhanes.red),]

dnhanes <- svydesign(id =~ sdmvpsu, strata =~ sdmvstra, weights =~ wtmec2yr, nest = TRUE, data = nhanes.red)

# Fit final model
finmod <- svyglm(bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat) + agec + agec2 +
  as.factor(ridreth1):agec + as.factor(ridreth1):agec2 + as.factor(riagendr):agec + as.factor(riagendr):agec2,
design = dnhanes)

# Replicate the final model results done above, note that degf is set to dnhanes for df in design
summary(finmod, df.resid = degf(dnhanes))

# Use svyCooksD for modified Cook's D
mcook <- svyCooksD(mobj = finmod, stvar = "sdmvstra", clvar = "sdmvpsu", doplot = TRUE)

# Examine cases > 10
mcook[mcook > 10]

#   2168     3459     5543     5904
# 18.78218 10.01648 10.47560 10.80141

# Create data frame without the 4 cases, examine refit model
nhanes1112a <- nhanes1112[-c(2168,3459,5543,5904),]

dnhanes2 <- svydesign(id =~ sdmvpsu, strata =~ sdmvstra, weights =~ wtmec2yr, nest = TRUE, data = nhanes1112a)

finmod2 <- svyglm(bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat) + agec + agec2 +
as.factor(ridreth1):agec + as.factor(ridreth1):agec2 + as.factor(riagendr):agec + as.factor(riagendr):agec2,
subset = (age18p == 1), design = dnhanes2)

summary(finmod2, df.resid = degf(dnhanes2)) # results indicate minor differences, see text for more information
about how to proceed


# Examine dfbetas to check influence of individual obs on parameter estimates

dfbetas <- svydfbetas(mobj = finmod, stvar = "sdmvstra", clvar = "sdmvpsu")
```

```
# Check influence of one case 2168
b <- data.frame(dfbetas$Dfbetas)
b$X2168

# [1] -0.019095264  0.015358191  0.017390378  0.016738493  0.018028044  0.006663257 -0.005607790
# [8] -0.001839896  0.061285194  0.065315288 -0.043712952 -0.040747920 -0.058707437 -0.037383482
# [15] -0.039585347 -0.075196788 -0.061617958 -0.058622136 -0.009626907 -0.009821412

# Collinearity diagnostics

# need numeric version of design matrix, for reduced CC data set
X <- model.matrix(~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat) +
agec + agec2 + as.factor(ridreth1):agec + as.factor(ridreth1):agec2 +
as.factor(riagendr):agec + as.factor(riagendr):agec2, data = nhanes.red)

svycollinear(mod=X, w=nhanes.red$wtmec2yr, Vcov=vcov(finmod), sc=TRUE, svyglm.obj=FALSE, rnd=3, fuzz=0.3)

# Investigate the variance inflation factors, largest are associated with age*age, not surprising
svyvif(mobj=finmod, X=X[,-1], w=nhanes.red$wtmec2yr, stvar = "sdmvstra", clvar = "sdmvpsu")

# Compute standardized residuals
st.resids <- svystdres(mobj = finmod, stvar = "sdmvstra", clvar = "sdmvpsu", doplot = TRUE)
# see automatically generated dot plot

st.resids$stdresids[st.resids$stdresids < -4]
# examine residuals < -4
#     2735      3732      3138      1158      5815
# -4.186686 -5.030542 -4.179745 -4.036181 -4.826700

# Re-run  Final Model including interactions of race X age, gender X age, and main effects of marital status
ex75_final_2 <- svyglm(bpxdi1_1 ~ marcatc + agec*factor(racec) + agecsq*factor(racec) + factor(genderc)*agec +
factor(genderc)*agecsq,
  subnhanes, family=gaussian)

# Check coefficients
(ex75_final_2$coefficients)

# Obtain summary of model
summary(ex75_final_2, df.resid=degf(subnhanes))

# Calculate CI as follows
#lowerci
coef(ex75_final_2) - (1.96*SE(ex75_final_2))

#upperci
coef(ex75_final_2) + (1.96*SE(ex75_final_2))


# 7.5.9 Q approach for weighting, see Pfefferman for details, table 7.6 column 2
# Use complete case data frame called nhanes.red from diagnostics analysis above

# Step 1 linear model with weight regressed on race, gender marcat and age / gender interactions
# In complete case adult data set >= 18 years of age

q_wgt <- lm(wtmec2yr ~ factor(marcat) + agec*factor(ridreth1) + agec2*factor(ridreth1) +
  factor(riagendr)*agec + factor(riagendr)*agec2 , nhanes.red)
summary(q_wgt)
```

```
w_hat <- predict(q_wgt)
summary(w_hat)

# Set values < 0 to 4809 (1 percentile value)
w_hat <- ifelse (w_hat < 0 , 4809,w_hat)
summary(w_hat)

# Create revised weight
nhanes.red$q_wtmec2yr = (nhanes.red$wtmec2yr / w_hat)

# Design object for analysis
nhanessvyq <- svydesign(strata=~sdmvstra, id=~sdmvpsu, weights=~q_wtmec2yr, data=nhanes.red, nest=T)

# Final model with revised Q weight, table 7.6
ex75_finalq <- svyglm(bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat) +
  agec + agec2 + as.factor(ridreth1):agec + as.factor(ridreth1):agec2 +
   as.factor(riagendr):agec + as.factor(riagendr):agec2, design = nhanessvyq)
summary(ex75_finalq, df=NULL,df.resid=Inf )


# Alternative Bayesian Approach

# load/install needed packages if not already in place
library(devtools)
#install_github("RyanHornby/csSampling") # lots of packages needed, will take a second
library(csSampling)
library(rstan)
library(brms)
library(survey)
rstan_options(auto_write = TRUE)

# Need to normalize NHANES weights to match what is done for Stan modeling
nhanes.red$wtsc <- nhanes.red$wtmec2yr / mean(nhanes.red$wtmec2yr)

# Survey design object
nhanes.des <- svydesign(id = ~sdmvpsu, strata = ~sdmvstra, weights = ~wtsc, nest = T, data = nhanes.red)

# Bayesian approach, flat prior
set.seed(41279)
model_formula <- formula("bpxdi1.1|weights(wtsc) ~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat)
+ agec + agec2 + as.factor(ridreth1):agec + as.factor(ridreth1):agec2 + as.factor(riagendr):agec +
as.factor(riagendr):agec2")
mod.brms <- cs_sampling_brms(svydes = nhanes.des, brmsmod = brmsformula(model_formula), data = nhanes.red,
family = gaussian())

# Print results
mod.brms$stan_fit
```

R Results

```
R version 4.4.1 (2024-06-14 ucrt) -- "Race for Your Life"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64


> # ASDA3 Analysis Examples for Chapter 7
>
> # Load packages and prepare data for analysis
> library (survey)
Loading required package: grid
Loading required package: Matrix
Loading required package: survival

Attaching package: ‘survey’

The following object is masked from ‘package:graphics’:

    dotchart

> library(haven)
>
> setwd("P:/ASDA3/Replication R/Chapter 7")
>
> load("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/nhanes1112.rdata")
> nhanesdata <- nhanes1112
>
> # Set key variables to values (0=NA for blood pressure, center age and create agecs2)
> nhanesdata$bpxdi1_1 <-nhanesdata$bpxdi1
> nhanesdata$bpxdi1_1 [nhanesdata$bpxdi1_1 ==0] <- NA
> nhanesdata$agec <- (nhanesdata$age-46.36)
> nhanesdata$agecsq <- (nhanesdata$agec*nhanesdata$agec)
> nhanesdata$genderc <- factor(nhanesdata$riagendr , levels = 1: 2, labels =c("M", "F"))
> summary(nhanesdata)
      seqn          ridstatr        riagendr        ridreth1        dmdmartl
 Min.   :62161   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   : 1.000
 1st Qu.:64600   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:3.000   1st Qu.: 1.000
 Median :67039   Median :2.000   Median :2.000   Median :3.000   Median : 2.000
 Mean   :67039   Mean   :1.957   Mean   :1.502   Mean   :3.229   Mean   : 2.749
 3rd Qu.:69477   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:4.000   3rd Qu.: 5.000
 Max.   :71916   Max.   :2.000   Max.   :2.000   Max.   :5.000   Max.   :99.000
                                                                 NA's   :4196
    wtint2yr         wtmec2yr         sdmvpsu         sdmvstra         indfmpir
 Min.   :  3321   Min.   :     0   Min.   :1.000   Min.   : 90.00   Min.   :0.000
 1st Qu.: 11352   1st Qu.: 11174   1st Qu.:1.000   1st Qu.: 92.00   1st Qu.:0.860
 Median : 18098   Median : 18090   Median :2.000   Median : 96.00   Median :1.630
 Mean   : 31426   Mean   : 31426   Mean   :1.643   Mean   : 95.87   Mean   :2.205
 3rd Qu.: 34887   3rd Qu.: 34792   3rd Qu.:2.000   3rd Qu.: 99.00   3rd Qu.:3.580
 Max.   :220233   Max.   :222580   Max.   :3.000   Max.   :103.00   Max.   :5.000
                                                                    NA's   :840
     bpxsy1          bpxdi1          bpxsy2          bpxdi2           bpxsy3
 Min.   : 74.0   Min.   :  0.0   Min.   : 74.0   Min.   :  0.00   Min.   : 74.0
 1st Qu.:106.0   1st Qu.: 60.0   1st Qu.:106.0   1st Qu.: 58.00   1st Qu.:106.0
 Median :116.0   Median : 68.0   Median :116.0   Median : 68.00   Median :116.0
 Mean   :119.2   Mean   : 66.9   Mean   :118.7   Mean   : 66.28   Mean   :118.2
 3rd Qu.:128.0   3rd Qu.: 76.0   3rd Qu.:128.0   3rd Qu.: 76.00   3rd Qu.:128.0
 Max.   :238.0   Max.   :120.0   Max.   :234.0   Max.   :134.00   Max.   :232.0
 NA's   :3000    NA's   :3000    NA's   :2848    NA's   :2848     NA's   :2839
```

```
     bpxdi3           bpxsy4         bpxdi4           bmxbmi          lbxtc
 Min.   :  0.00   Min.   : 78    Min.   :  0.00   Min.   :12.40   Min.   : 59.0
 1st Qu.: 58.00   1st Qu.:104    1st Qu.: 64.00   1st Qu.:19.30   1st Qu.:154.0
 Median : 68.00   Median :116    Median : 72.00   Median :24.50   Median :179.0
 Mean   : 65.91   Mean   :119    Mean   : 71.78   Mean   :25.34   Mean   :183.2
 3rd Qu.: 76.00   3rd Qu.:130    3rd Qu.: 82.00   3rd Qu.:29.80   3rd Qu.:209.0
 Max.   :128.00   Max.   :226    Max.   :130.00   Max.   :82.10   Max.   :523.0
 NA's   :2839     NA's   :9309   NA's   :9309     NA's   :1154    NA's   :2768
     age18p         irregular          edcat            age           marcat
 Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   : 0.0   Min.   :1.000
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.: 9.0   1st Qu.:1.000
 Median :1.0000   Median :0.0000   Median :2.000   Median :26.0   Median :1.000
 Mean   :0.6011   Mean   :0.0134   Mean   :2.102   Mean   :31.4   Mean   :1.652
 3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:3.000   3rd Qu.:52.0   3rd Qu.:2.000
 Max.   :1.0000   Max.   :1.0000   Max.   :4.000   Max.   :80.0   Max.   :3.000
                  NA's   :729      NA's   :1602                   NA's   :4203
    pre_hibp          bp_cat          ag1829          ag3044          ag4559
 Min.   :0.0000   Min.   :1.000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
 1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
 Median :0.0000   Median :1.000   Median :1.000   Median :0.0000   Median :0.0000
 Mean   :0.4357   Mean   :1.546   Mean   :0.532   Mean   :0.1461   Mean   :0.1384
 3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:0.0000
 Max.   :1.0000   Max.   :4.000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
 NA's   :2701     NA's   :2701
      ag60             mex             othhis           white            black
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
 Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000   Median :0.000
 Mean   :0.1836   Mean   :0.1389   Mean   :0.1103   Mean   :0.3047   Mean   :0.275
 3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000

     other            lbdhdd          lbdhddsi        lbdtcsi          bpxdi1_1
 Min.   :0.0000   Min.   : 14.00   Min.   :0.360   Min.   : 1.530   Min.   : 10.0
 1st Qu.:0.0000   1st Qu.: 43.00   1st Qu.:1.110   1st Qu.: 3.980   1st Qu.: 60.0
 Median :0.0000   Median : 51.00   Median :1.320   Median : 4.630   Median : 68.0
 Mean   :0.1711   Mean   : 52.63   Mean   :1.361   Mean   : 4.738   Mean   : 67.7
 3rd Qu.:0.0000   3rd Qu.: 60.00   3rd Qu.:1.550   3rd Qu.: 5.400   3rd Qu.: 76.0
 Max.   :1.0000   Max.   :175.00   Max.   :4.530   Max.   :13.520   Max.   :120.0
                  NA's   :2767     NA's   :2767    NA's   :2768     NA's   :3080
      agec             agecsq          genderc
 Min.   :-46.36   Min.   :   0.1296   M:4856
 1st Qu.:-37.36   1st Qu.: 178.4896   F:4900
 Median :-20.36   Median : 709.6896
 Mean   :-14.96   Mean   : 827.7879
 3rd Qu.:  5.64   3rd Qu.:1395.7696
 Max.   : 33.64   Max.   :2149.2496
```

```
> # Create factor variables for models
> nhanesdata$racec <- factor(nhanesdata$ridreth1, levels = 1: 5 , labels =c("Mexican", "Other Hispanic",
"White",
+ "Black", "Other"))
> nhanesdata$marcatc <- factor(nhanesdata$marcat, levels = 1: 3, labels =c("Married", "Previously Married",
"Never Married"))
> nhanesdata$edcatc <- factor(nhanesdata$edcat, levels = 1: 4, labels =c("0-11", "12", "13-15","16+"))
> nhanesdata$bp_catc <- factor(nhanesdata$bp_cat, levels = 1: 4, labels =c("Normal", "Pre-HBP", "Stage 1
HBP","Stage 2 HBP"))
```

```
> # Subpop indicator
> nhanesdata$age18p <- ifelse(nhanesdata$age >=18,1,0)
>
> # Survey design
> nhanessvy2 <- svydesign(strata=~sdmvstra, id=~sdmvpsu, weights=~wtmec2yr, data=nhanesdata, nest=T)
> subnhanes <- subset(nhanessvy2 , age >= 18)
>
>
> # Table 7.1 Numbers from Bivariate Tests of Race, Marital Status, Gender, Centered Age
> # Race/ethnicity
> summary(ex75_race <- svyglm(bpxdi1_1 ~racec, design=subnhanes))

Call:
svyglm(formula = bpxdi1_1 ~ racec, design = subnhanes)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          69.8041     0.4532 154.013  < 2e-16 ***
racecOther Hispanic  -0.1549     1.4556  -0.106  0.91688
racecWhite            2.1847     0.7427   2.942  0.01145 *
racecBlack            2.2902     0.7030   3.258  0.00623 **
racecOther            1.3056     0.7044   1.853  0.08665 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 137.0384)

Number of Fisher Scoring iterations: 2

> regTermTest(ex75_race, ~racec)
Wald test for racec
 in svyglm(formula = bpxdi1_1 ~ racec, design = subnhanes)
F =  4.771214  on  4  and  13  df: p= 0.013705
>
> # Marital Status
> summary(ex75_marital <- svyglm(bpxdi1_1 ~marcatc, design=subnhanes))

Call:
svyglm(formula = bpxdi1_1 ~ marcatc, design = subnhanes)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                72.1796     0.5149 140.172   <2e-16 ***
marcatcPreviously Married  -0.1451     0.6978  -0.208    0.838
marcatcNever Married       -1.1210     0.8437  -1.329    0.204
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 137.5607)

Number of Fisher Scoring iterations: 2

> regTermTest(ex75_marital, ~marcatc)
```

```
Wald test for marcatc
 in svyglm(formula = bpxdi1_1 ~ marcatc, design = subnhanes)
F =  0.9023684  on  2  and  15  df: p= 0.42653
>
> # Gender
> summary(ex75_sex <- svyglm(bpxdi1_1 ~genderc, design=subnhanes))

Call:
svyglm(formula = bpxdi1_1 ~ genderc, design = subnhanes)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  72.7255     0.5901 123.245  < 2e-16 ***
gendercF     -2.2004     0.5679  -3.875  0.00134 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 136.4476)

Number of Fisher Scoring iterations: 2

> regTermTest(ex75_sex, ~genderc)
Wald test for genderc
 in svyglm(formula = bpxdi1_1 ~ genderc, design = subnhanes)
F =  15.01184  on  1  and  16  df: p= 0.0013441
>
> # Centered age
> summary(ex75_age <- svyglm(bpxdi1_1 ~agec, design=subnhanes))

Call:
svyglm(formula = bpxdi1_1 ~ agec, design = subnhanes)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 71.60363    0.50024 143.140   <2e-16 ***
agec         0.03941    0.01889   2.087   0.0533 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 137.2234)

Number of Fisher Scoring iterations: 2
```

```
> # 7.5.2 naive analysis, Table 7.2
> (ex75_nowt <- lm(bpxdi1_1 ~ racec + marcatc + genderc + agec, data=nhanesdata, age >=18 ))

Call:
lm(formula = bpxdi1_1 ~ racec + marcatc + genderc + agec, data = nhanesdata,
    subset = age >= 18)

Coefficients:
              (Intercept)        racecOther Hispanic                 racecWhite
                 72.16282                    0.37705                    0.94207
                racecBlack                racecOther  marcatcPreviously Married
                  2.50903                    1.77582                   -0.69731
     marcatcNever Married                  gendercF                       agec
                 -2.44035                   -2.53233                   -0.01925

> summary(ex75_nowt)

Call:
lm(formula = bpxdi1_1 ~ racec + marcatc + genderc + agec, data = nhanesdata,
    subset = age >= 18)

Residuals:
    Min      1Q  Median      3Q     Max
-59.228  -7.233   0.022   7.297  48.012

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               72.16282    0.56762 127.133  < 2e-16 ***
racecOther Hispanic        0.37705    0.75603   0.499  0.61800
racecWhite                 0.94207    0.61036   1.543  0.12278
racecBlack                 2.50903    0.63979   3.922 8.92e-05 ***
racecOther                 1.77582    0.67721   2.622  0.00876 **
marcatcPreviously Married -0.69731    0.44364  -1.572  0.11607
marcatcNever Married      -2.44035    0.46228  -5.279 1.36e-07 ***
gendercF                  -2.53233    0.33925  -7.464 9.87e-14 ***
agec                      -0.01925    0.01087  -1.770  0.07673 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.66 on 4836 degrees of freedom
  (1019 observations deleted due to missingness)
Multiple R-squared:  0.02071,   Adjusted R-squared:  0.01909
F-statistic: 12.78 on 8 and 4836 DF,  p-value: < 2.2e-16
```

```
> # 7.5.3 weighted linear regression without design variables, Table 7.3
> (ex75_wt <- lm(bpxdi1_1 ~ racec + marcatc + genderc + agec, data= nhanesdata, age >=18, weight=wtmec2yr ))

Call:
lm(formula = bpxdi1_1 ~ racec + marcatc + genderc + agec, data = nhanesdata,
    subset = age >= 18, weights = wtmec2yr)

Coefficients:
              (Intercept)        racecOther Hispanic                 racecWhite
                71.806173                  -0.094977                   1.913279
                racecBlack                 racecOther  marcatcPreviously Married
                 2.619169                   1.270136                    0.243307
     marcatcNever Married                   gendercF                        agec
                -1.511033                  -2.509811                   -0.007537

> summary(ex75_wt)

Call:
lm(formula = bpxdi1_1 ~ racec + marcatc + genderc + agec, data = nhanesdata,
    subset = age >= 18, weights = wtmec2yr)

Weighted Residuals:
   Min     1Q Median     3Q    Max
-16985  -1286    -27   1132  14770

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               71.806173   0.613851 116.977  < 2e-16 ***
racecOther Hispanic       -0.094977   0.864694  -0.110 0.912542
racecWhite                 1.913279   0.625420   3.059 0.002232 **
racecBlack                 2.619169   0.764390   3.426 0.000617 ***
racecOther                 1.270136   0.827338   1.535 0.124798
marcatcPreviously Married  0.243307   0.442626   0.550 0.582558
marcatcNever Married      -1.511033   0.454074  -3.328 0.000882 ***
gendercF                  -2.509811   0.327532  -7.663 2.18e-14 ***
agec                      -0.007537   0.010892  -0.692 0.489024
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2316 on 4836 degrees of freedom
  (1019 observations deleted due to missingness)
Multiple R-squared:  0.01745,   Adjusted R-squared:  0.01582
F-statistic: 10.74 on 8 and 4836 DF,  p-value: 4.393e-15

>
>
```

```
> #  7.5.4 incorporating all design features, Table 7.4
> summary(ex75_svyglm <- svyglm(bpxdi1_1 ~ racec + marcatc + genderc + agec, deff=T, design=subnhanes))

Call:
svyglm(formula = bpxdi1_1 ~ racec + marcatc + genderc + agec,
    design = subnhanes, deff = T)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               71.806173   0.577886 124.257 7.19e-16 ***
racecOther Hispanic       -0.094977   1.272263  -0.075  0.94212
racecWhite                 1.913279   0.762891   2.508  0.03342 *
racecBlack                 2.619169   0.599976   4.365  0.00181 **
racecOther                 1.270136   0.604242   2.102  0.06490 .
marcatcPreviously Married  0.243307   0.672400   0.362  0.72582
marcatcNever Married      -1.511033   0.899370  -1.680  0.12724
gendercF                  -2.509811   0.539294  -4.654  0.00120 **
agec                      -0.007537   0.020091  -0.375  0.71626
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 135.3628)

Number of Fisher Scoring iterations: 2

> deff(ex75_svyglm) # note that these differ from Stata, see documentation for details
              (Intercept)       racecOther Hispanic                racecWhite
                0.8862543                 2.1648574                 1.4879257
               racecBlack                racecOther marcatcPreviously Married
                0.6160822                 0.5334029                 2.3077181
     marcatcNever Married                  gendercF                      agec
                3.9230510                 2.7111009                 3.4023228
> plot(ex75_svyglm) # this produces many plots, sift through for plots of interest in plot window
Waiting to confirm page change...
Waiting to confirm page change...
Waiting to confirm page change...
Waiting to confirm page change...
#Not included here, see Graphics window in R when running code!


> nhanessvy_nowt <- svydesign(strata=~sdmvstra, id=~sdmvpsu, data=nhanesdata, nest=T)
Warning message:
In svydesign.default(strata = ~sdmvstra, id = ~sdmvpsu, data = nhanesdata,  :
  No weights or probabilities supplied, assuming equal probability
> subnhanes_nowt <- subset(nhanessvy_nowt , age >= 18)
>
> # 7.5.4  Run model and do test of weight and weight interactions
> summary(ex75_nowt_svyglm <- svyglm(bpxdi1_1 ~ wtmec2yr + racec + genderc + agec +
+ wtmec2yr:(racec) + wtmec2yr:(genderc) + wtmec2yr:(marcatc) + wtmec2yr:agec , deff=T, design=subnhanes_nowt))

Call:
svyglm(formula = bpxdi1_1 ~ wtmec2yr + racec + genderc + agec +
    wtmec2yr:(racec) + wtmec2yr:(genderc) + wtmec2yr:(marcatc) +
    wtmec2yr:agec, design = subnhanes_nowt, deff = T)

Survey design:
```

```
subset(nhanessvy_nowt, age >= 18)

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                        7.243e+01  1.966e+00  36.851 0.000736 ***
wtmec2yr                          -1.487e-05  4.751e-05  -0.313 0.783986
racecOther Hispanic                1.353e+00  2.877e+00   0.470 0.684374
racecWhite                        -2.496e+00  1.763e+00  -1.416 0.292458
racecBlack                        -5.119e-01  2.347e+00  -0.218 0.847562
racecOther                         1.708e+00  1.957e+00   0.873 0.474758
gendercF                          -2.943e+00  4.570e-01  -6.439 0.023278 *
agec                              -2.034e-03  2.345e-02  -0.087 0.938767
wtmec2yr:racecOther Hispanic      -4.719e-05  5.415e-05  -0.871 0.475379
wtmec2yr:racecWhite                4.945e-05  4.685e-05   1.056 0.401802
wtmec2yr:racecBlack                1.239e-04  6.923e-05   1.789 0.215503
wtmec2yr:racecOther               -1.811e-05  5.251e-05  -0.345 0.763108
wtmec2yr:gendercF                  3.673e-06  6.899e-06   0.532 0.647646
wtmec2yr:marcatcPreviously Married 8.439e-06  8.534e-06   0.989 0.426942
wtmec2yr:marcatcNever Married     -1.755e-05  1.180e-05  -1.487 0.275450
wtmec2yr:agec                      1.502e-08  3.183e-07   0.047 0.966657
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 134.8929)

Number of Fisher Scoring iterations: 2

> coef(ex75_nowt_svyglm)
                       (Intercept)                            wtmec2yr
                      7.243436e+01                       -1.486553e-05
               racecOther Hispanic                          racecWhite
                      1.353167e+00                       -2.495709e+00
                        racecBlack                          racecOther
                     -5.118708e-01                        1.707877e+00
                           gendercF                                agec
                     -2.942554e+00                       -2.034128e-03
      wtmec2yr:racecOther Hispanic                 wtmec2yr:racecWhite
                     -4.719082e-05                        4.945465e-05
               wtmec2yr:racecBlack                 wtmec2yr:racecOther
                      1.238542e-04                       -1.810673e-05
                 wtmec2yr:gendercF wtmec2yr:marcatcPreviously Married
                      3.673361e-06                        8.439091e-06
     wtmec2yr:marcatcNever Married                       wtmec2yr:agec
                     -1.755032e-05                        1.501696e-08
> summary(ex75_nowt_svyglm)

Call:
svyglm(formula = bpxdi1_1 ~ wtmec2yr + racec + genderc + agec +
    wtmec2yr:(racec) + wtmec2yr:(genderc) + wtmec2yr:(marcatc) +
    wtmec2yr:agec, design = subnhanes_nowt, deff = T)

Survey design:
subset(nhanessvy_nowt, age >= 18)

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                        7.243e+01  1.966e+00  36.851 0.000736 ***
wtmec2yr                          -1.487e-05  4.751e-05  -0.313 0.783986
racecOther Hispanic                1.353e+00  2.877e+00   0.470 0.684374
```

```
racecWhite                          -2.496e+00  1.763e+00  -1.416 0.292458
racecBlack                          -5.119e-01  2.347e+00  -0.218 0.847562
racecOther                           1.708e+00  1.957e+00   0.873 0.474758
gendercF                            -2.943e+00  4.570e-01  -6.439 0.023278 *
agec                                -2.034e-03  2.345e-02  -0.087 0.938767
wtmec2yr:racecOther Hispanic        -4.719e-05  5.415e-05  -0.871 0.475379
wtmec2yr:racecWhite                  4.945e-05  4.685e-05   1.056 0.401802
wtmec2yr:racecBlack                  1.239e-04  6.923e-05   1.789 0.215503
wtmec2yr:racecOther                 -1.811e-05  5.251e-05  -0.345 0.763108
wtmec2yr:gendercF                    3.673e-06  6.899e-06   0.532 0.647646
wtmec2yr:marcatcPreviously Married   8.439e-06  8.534e-06   0.989 0.426942
wtmec2yr:marcatcNever Married       -1.755e-05  1.180e-05  -1.487 0.275450
wtmec2yr:agec                        1.502e-08  3.183e-07   0.047 0.966657
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 134.8929)

Number of Fisher Scoring iterations: 2


> # Note: df differs from Stata but the set of variables p value still shows significant as a group, weights are
informative
> regTermTest(ex75_nowt_svyglm, ~wtmec2yr + wtmec2yr:agec + wtmec2yr:racec + wtmec2yr:genderc + wtmec2yr:marcatc
, df=NULL)
Wald test for wtmec2yr wtmec2yr:agec wtmec2yr:racec wtmec2yr:genderc wtmec2yr:marcatc
 in svyglm(formula = bpxdi1_1 ~ wtmec2yr + racec + genderc + agec +
    wtmec2yr:(racec) + wtmec2yr:(genderc) + wtmec2yr:(marcatc) +
    wtmec2yr:agec, design = subnhanes_nowt, deff = T)
F =  27.58175  on  9  and  2  df: p= 0.035466
```

```
> # Example 7.5 re-run initial model with informative weights and assess basic regression diagnostics
> summary(ex75ini_svyglm <- svyglm(bpxdi1_1 ~ racec + marcatc + genderc + agec, deff=T, design=subnhanes))

Call:
svyglm(formula = bpxdi1_1 ~ racec + marcatc + genderc + agec,
    design = subnhanes, deff = T)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               71.806173   0.577886 124.257 7.19e-16 ***
racecOther Hispanic       -0.094977   1.272263  -0.075  0.94212
racecWhite                 1.913279   0.762891   2.508  0.03342 *
racecBlack                 2.619169   0.599976   4.365  0.00181 **
racecOther                 1.270136   0.604242   2.102  0.06490 .
marcatcPreviously Married  0.243307   0.672400   0.362  0.72582
marcatcNever Married      -1.511033   0.899370  -1.680  0.12724
gendercF                  -2.509811   0.539294  -4.654  0.00120 **
agec                      -0.007537   0.020091  -0.375  0.71626
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 135.3628)

Number of Fisher Scoring iterations: 2

> deff(ex75ini_svyglm)
              (Intercept)       racecOther Hispanic                racecWhite
                0.8862543                 2.1648574                 1.4879257
               racecBlack                racecOther marcatcPreviously Married
                0.6160822                 0.5334029                 2.3077181
     marcatcNever Married                  gendercF                      agec
                3.9230510                 2.7111009                 3.4023228
>
> plot(ex75ini_svyglm) # basic plots, see documentation for various other plotting options available in R from
svy commands
Waiting to confirm page change...
Waiting to confirm page change...
Waiting to confirm page change...
Waiting to confirm page change...
>
```

```
> # Example 7.5 with agecsq (agec squared), Table 7.5
> summary(ex75_svyglm_agesq <- svyglm(bpxdi1_1 ~ racec + marcatc + genderc + agec + agecsq, deff=T,
design=subnhanes))

Call:
svyglm(formula = bpxdi1_1 ~ racec + marcatc + genderc + agec +
    agecsq, design = subnhanes, deff = T)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               74.4613657  0.6151182 121.052 2.42e-14 ***
racecOther Hispanic        0.0145334  1.1137793   0.013  0.98991
racecWhite                 1.8217232  0.7929802   2.297  0.05068 .
racecBlack                 2.0635973  0.6120855   3.371  0.00977 **
racecOther                 1.0978333  0.5839260   1.880  0.09689 .
marcatcPreviously Married  0.5558885  0.6076709   0.915  0.38705
marcatcNever Married       1.1353283  0.7577263   1.498  0.17243
gendercF                  -2.2143152  0.5093059  -4.348  0.00245 **
agec                       0.0771555  0.0204888   3.766  0.00550 **
agecsq                    -0.0116650  0.0008549 -13.645 8.01e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 124.2041)

Number of Fisher Scoring iterations: 2

> summary(ex75_svyglm_agesq)

Call:
svyglm(formula = bpxdi1_1 ~ racec + marcatc + genderc + agec +
    agecsq, design = subnhanes, deff = T)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               74.4613657  0.6151182 121.052 2.42e-14 ***
racecOther Hispanic        0.0145334  1.1137793   0.013  0.98991
racecWhite                 1.8217232  0.7929802   2.297  0.05068 .
racecBlack                 2.0635973  0.6120855   3.371  0.00977 **
racecOther                 1.0978333  0.5839260   1.880  0.09689 .
marcatcPreviously Married  0.5558885  0.6076709   0.915  0.38705
marcatcNever Married       1.1353283  0.7577263   1.498  0.17243
gendercF                  -2.2143152  0.5093059  -4.348  0.00245 **
agec                       0.0771555  0.0204888   3.766  0.00550 **
agecsq                    -0.0116650  0.0008549 -13.645 8.01e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 124.2041)

Number of Fisher Scoring iterations: 2

>
```

```
> # Preliminary final model with race X age interactions
> ex75_raceint <- svyglm(bpxdi1_1 ~ marcatc + genderc + factor(racec)*agec + factor(racec)*agecsq , subnhanes)
> summary(ex75_raceint, df.resid=Inf)

Call:
svyglm(formula = bpxdi1_1 ~ marcatc + genderc + factor(racec) *
    agec + factor(racec) * agecsq, design = subnhanes)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                        7.445e+01  8.429e-01  88.328  < 2e-16 ***
marcatcPreviously Married          5.488e-01  6.295e-01   0.872 0.383279
marcatcNever Married               1.188e+00  7.784e-01   1.526 0.126898
gendercF                          -2.214e+00  5.052e-01  -4.383 1.17e-05 ***
factor(racec)Other Hispanic        3.457e-01  9.839e-01   0.351 0.725306
factor(racec)White                 1.552e+00  9.393e-01   1.653 0.098374 .
factor(racec)Black                 3.394e+00  9.911e-01   3.424 0.000617 ***
factor(racec)Other                 1.380e+00  8.781e-01   1.572 0.116062
agec                               5.024e-02  3.501e-02   1.435 0.151249
agecsq                            -1.245e-02  2.210e-03  -5.631 1.79e-08 ***
factor(racec)Other Hispanic:agec   6.626e-02  5.026e-02   1.318 0.187378
factor(racec)White:agec            1.112e-02  5.394e-02   0.206 0.836691
factor(racec)Black:agec            6.940e-02  3.704e-02   1.873 0.061011 .
factor(racec)Other:agec            4.385e-02  5.902e-02   0.743 0.457548
factor(racec)Other Hispanic:agecsq 5.103e-05  3.350e-03   0.015 0.987847
factor(racec)White:agecsq          1.850e-03  1.791e-03   1.033 0.301773
factor(racec)Black:agecsq         -3.912e-03  2.209e-03  -1.771 0.076576 .
factor(racec)Other:agecsq         -1.547e-04  2.248e-03  -0.069 0.945109
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero or negative residual df; p-values not defined

(Dispersion parameter for gaussian family taken to be 123.8154)

Number of Fisher Scoring iterations: 2

>
> # Joint test of race X age interactions
> regTermTest(ex75_raceint, ~ factor(racec):agec + factor(racec):agecsq, df=Inf)
Wald test for factor(racec):agec factor(racec):agecsq
 in svyglm(formula = bpxdi1_1 ~ marcatc + genderc + factor(racec) *
    agec + factor(racec) * agecsq, design = subnhanes)
Chisq =  80.51104  on  8  df: p= 3.8576e-14
```

```
> # Preliminary final model plus age vars X gender interactions
> ex75_genint <- svyglm(bpxdi1_1 ~ marcatc + racec + factor(genderc)*agec + factor(genderc)*agecsq , subnhanes)
> summary(ex75_genint, df.resid=Inf)

Call:
svyglm(formula = bpxdi1_1 ~ marcatc + racec + factor(genderc) *
    agec + factor(genderc) * agecsq, design = subnhanes)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               74.9687964  0.7033104 106.594  < 2e-16 ***
marcatcPreviously Married  0.2931299  0.6195993   0.473  0.63614
marcatcNever Married       1.0564049  0.7546553   1.400  0.16156
racecOther Hispanic        0.0164345  1.1032920   0.015  0.98812
racecWhite                 1.8370567  0.7811790   2.352  0.01869 *
racecBlack                 2.1249157  0.6122988   3.470  0.00052 ***
racecOther                 1.1144157  0.5913832   1.884  0.05951 .
factor(genderc)F          -3.0823607  0.7562744  -4.076 4.59e-05 ***
agec                       0.0504735  0.0208039   2.426  0.01526 *
agecsq                    -0.0133454  0.0009431 -14.151  < 2e-16 ***
factor(genderc)F:agec      0.0503281  0.0255313   1.971  0.04870 *
factor(genderc)F:agecsq    0.0029326  0.0016936   1.732  0.08335 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 123.7001)

Number of Fisher Scoring iterations: 2


> # Joint test of gender X age interactions
> regTermTest(ex75_genint, ~ factor(genderc):agec + factor(genderc):agecsq, df=Inf, null=NULL)
Wald test for factor(genderc):agec factor(genderc):agecsq
 in svyglm(formula = bpxdi1_1 ~ marcatc + racec + factor(genderc) *
    agec + factor(genderc) * agecsq, design = subnhanes)
Chisq =  9.072366  on  2  df: p= 0.010714
```

```
> # Final model including interactions of race X age, gender X age
> ex75_final <- svyglm(bpxdi1_1 ~ marcatc + agec*factor(racec) + agecsq*factor(racec) + factor(genderc)*agec +
+  factor(genderc)*agecsq, subnhanes)
> summary(ex75_final, df.resid=Inf)

Call:
svyglm(formula = bpxdi1_1 ~ marcatc + agec * factor(racec) +
    agecsq * factor(racec) + factor(genderc) * agec + factor(genderc) *
    agecsq, design = subnhanes)

Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                     74.9279514  0.9327484  80.330  < 2e-16 ***
marcatcPreviously Married        0.2905908  0.6398622   0.454 0.649724
marcatcNever Married             1.1076855  0.7757962   1.428 0.153348
agec                             0.0281912  0.0397675   0.709 0.478386
factor(racec)Other Hispanic      0.4060606  0.9797099   0.414 0.678530
factor(racec)White               1.6140818  0.9440749   1.710 0.087322 .
factor(racec)Black               3.5198661  0.9909619   3.552 0.000382 ***
factor(racec)Other               1.4164313  0.8617961   1.644 0.100263
agecsq                          -0.0138641  0.0023774  -5.832 5.49e-09 ***
factor(genderc)F                -3.1027738  0.7561240  -4.104 4.07e-05 ***
agec:factor(racec)Other Hispanic 0.0605772  0.0519876   1.165 0.243929
agec:factor(racec)White          0.0076904  0.0564603   0.136 0.891657
agec:factor(racec)Black          0.0646497  0.0391346   1.652 0.098538 .
agec:factor(racec)Other          0.0390183  0.0620544   0.629 0.529496
factor(racec)Other Hispanic:agecsq -0.0003224  0.0036160  -0.089 0.928963
factor(racec)White:agecsq        0.0015388  0.0019529   0.788 0.430717
factor(racec)Black:agecsq       -0.0042949  0.0023828  -1.802 0.071477 .
factor(racec)Other:agecsq       -0.0003766  0.0024760  -0.152 0.879094
agec:factor(genderc)F            0.0484726  0.0259984   1.864 0.062259 .
agecsq:factor(genderc)F          0.0030105  0.0016991   1.772 0.076432 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero or negative residual df; p-values not defined

(Dispersion parameter for gaussian family taken to be 123.3161)

Number of Fisher Scoring iterations: 2
```

 # MARGINAL EFFECTS EXAMPLES ARE FORTHCOMING, CHECK WEBSITE FOR UPDATES!

```
> # Model diagnostics with survey features incorporated
> # Load svydiags package
> library(svydiags)
Loading required package: MASS
Warning message:
package ‘MASS’ was built under R version 4.4.3


> # Note the variables names are slightly different than models previously run but are equivalent:
bpxdi1.1=bpxdi1_1 etc.

> load("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/nhanes1112.rdata")
>
>  nhanes1112$bpxdi1.1 <- nhanes1112$bpxdi1
>  nhanes1112$bpxdi1.1[nhanes1112$bpxdi1 == 0] <- NA
>  nhanes1112$agec <- nhanes1112$age - 46.36
>  nhanes1112$agec2 <- nhanes1112$agec ^ 2
>
> # Select needed variables and obtain complete case frame
> nhanes.red <- nhanes1112[, c("bpxdi1.1", "ridreth1", "riagendr", "marcat", "agec", "agec2", "age18p",
"wtmec2yr", "sdmvstra", "sdmvpsu")]
> nhanes.red <- nhanes.red[complete.cases(nhanes.red),]
>
> dnhanes <- svydesign(id =~ sdmvpsu, strata =~ sdmvstra, weights =~ wtmec2yr, nest = TRUE, data = nhanes.red)
>
>
> # Fit final model
> finmod <- svyglm(bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat) + agec + agec2 +
+   as.factor(ridreth1):agec + as.factor(ridreth1):agec2 + as.factor(riagendr):agec + as.factor(riagendr):agec2,
design = dnhanes)
>
> # Replicate the final model results done above, note that degf is set to dnhanes for df in design
> summary(finmod, df.resid = degf(dnhanes))

Call:
svyglm(formula = bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) +
    as.factor(marcat) + agec + agec2 + as.factor(ridreth1):agec +
    as.factor(ridreth1):agec2 + as.factor(riagendr):agec + as.factor(riagendr):agec2,
    design = dnhanes)

Survey design:
svydesign(id = ~sdmvpsu, strata = ~sdmvstra, weights = ~wtmec2yr,
    nest = TRUE, data = nhanes.red)

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               74.9279514  0.9327484  80.330  < 2e-16 ***
as.factor(ridreth1)2       0.4060606  0.9797099   0.414 0.683713
as.factor(ridreth1)3       1.6140818  0.9440749   1.710 0.105510
as.factor(ridreth1)4       3.5198661  0.9909619   3.552 0.002452 **
as.factor(ridreth1)5       1.4164313  0.8617961   1.644 0.118628
as.factor(riagendr)2      -3.1027738  0.7561240  -4.104 0.000741 ***
as.factor(marcat)2         0.2905908  0.6398622   0.454 0.655469
as.factor(marcat)3         1.1076855  0.7757962   1.428 0.171461
agec                       0.0281912  0.0397675   0.709 0.487995
agec2                     -0.0138641  0.0023774  -5.832    2e-05 ***
as.factor(ridreth1)2:agec  0.0605772  0.0519876   1.165 0.260019
as.factor(ridreth1)3:agec  0.0076904  0.0564603   0.136 0.893257
as.factor(ridreth1)4:agec  0.0646497  0.0391346   1.652 0.116887
```

```
as.factor(ridreth1)5:agec   0.0390183  0.0620544   0.629 0.537853
as.factor(ridreth1)2:agec2 -0.0003224  0.0036160  -0.089 0.930005
as.factor(ridreth1)3:agec2  0.0015388  0.0019529   0.788 0.441566
as.factor(ridreth1)4:agec2 -0.0042949  0.0023828  -1.802 0.089235 .
as.factor(ridreth1)5:agec2 -0.0003766  0.0024760  -0.152 0.880885
as.factor(riagendr)2:agec   0.0484726  0.0259984   1.864 0.079620 .
as.factor(riagendr)2:agec2  0.0030105  0.0016991   1.772 0.094355 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero or negative residual df; p-values not defined

(Dispersion parameter for gaussian family taken to be 115.0377)

Number of Fisher Scoring iterations: 2


> # Use svyCooksD for modified Cook's D
> mcook <- svyCooksD(mobj = finmod, stvar = "sdmvstra", clvar = "sdmvpsu", doplot = TRUE)


> # Examine cases > 10
> mcook[mcook > 10]
    2168     3459     5543     5904
18.78218 10.01648 10.47560 10.80141


> # Create data frame without the 4 cases, examine refit model
> nhanes1112a <- nhanes1112[-c(2168,3459,5543,5904),]

> dnhanes2 <- svydesign(id =~ sdmvpsu, strata =~ sdmvstra, weights =~ wtmec2yr, nest = TRUE, data = nhanes1112a)

> finmod2 <- svyglm(bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat) + agec + agec2 +
as.factor(ridreth1):agec + as.factor(ridreth1):agec2 + as.factor(riagendr):agec + as.factor(riagendr):agec2,
subset = (age18p == 1), design = dnhanes2)

> summary(finmod2, df.resid = degf(dnhanes2)) # results indicate minor differences, see text for more
information about how to proceed

Call:
svyglm(formula = bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) +
    as.factor(marcat) + agec + agec2 + as.factor(ridreth1):agec +
    as.factor(ridreth1):agec2 + as.factor(riagendr):agec + as.factor(riagendr):agec2,
    design = dnhanes2, subset = (age18p == 1))

Survey design:
svydesign(id = ~sdmvpsu, strata = ~sdmvstra, weights = ~wtmec2yr,
    nest = TRUE, data = nhanes1112a)

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               74.9090490  0.9411027  79.597  < 2e-16 ***
as.factor(ridreth1)2       0.4274829  0.9783597   0.437  0.66766
as.factor(ridreth1)3       1.6361082  0.9349400   1.750  0.09815 .
as.factor(ridreth1)4       3.5417417  0.9883623   3.583  0.00229 **
as.factor(ridreth1)5       1.4388676  0.8627622   1.668  0.11368
as.factor(riagendr)2      -3.1106896  0.7555964  -4.117  0.00072 ***
as.factor(marcat)2         0.2974973  0.6396387   0.465  0.64776
as.factor(marcat)3         1.1088024  0.7746555   1.431  0.17046
```

```
agec                         0.0301554  0.0362326   0.832  0.41679
agec2                       -0.0136610  0.0022835  -5.982 1.48e-05 ***
as.factor(ridreth1)2:agec    0.0582728  0.0498529   1.169  0.25858
as.factor(ridreth1)3:agec    0.0054483  0.0525053   0.104  0.91857
as.factor(ridreth1)4:agec    0.0623529  0.0351376   1.775  0.09388 .
as.factor(ridreth1)5:agec    0.0367827  0.0597118   0.616  0.54605
as.factor(ridreth1)2:agec2 -0.0005500  0.0035036  -0.157  0.87711
as.factor(ridreth1)3:agec2  0.0013119  0.0018825   0.697  0.49527
as.factor(ridreth1)4:agec2 -0.0045238  0.0022778  -1.986  0.06339 .
as.factor(ridreth1)5:agec2 -0.0006042  0.0024245  -0.249  0.80618
as.factor(riagendr)2:agec    0.0489156  0.0261060   1.874  0.07826 .
as.factor(riagendr)2:agec2  0.0030513  0.0016927   1.803  0.08921 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero or negative residual df; p-values not defined

(Dispersion parameter for gaussian family taken to be 123.3192)

Number of Fisher Scoring iterations: 2


> # Examine dfbetas to check influence of individual obs on parameter estimates

> dfbetas <- svydfbetas(mobj = finmod, stvar = "sdmvstra", clvar = "sdmvpsu")

> # Check influence of one case 2168
> b <- data.frame(dfbetas$Dfbetas)
> b$X2168
 [1] -0.019095264  0.015358191  0.017390378  0.016738493  0.018028044  0.006663257
 [7] -0.005607790 -0.001839896  0.061285194  0.065315288 -0.043712952 -0.040747920
[13] -0.058707437 -0.037383482 -0.039585347 -0.075196788 -0.061617958 -0.058622136
[19] -0.009626907 -0.009821412

> # Collinearity diagnostics

> # need numeric version of design matrix, for reduced CC data set
> X <- model.matrix(~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat) +
+ agec + agec2 + as.factor(ridreth1):agec + as.factor(ridreth1):agec2 +
+ as.factor(riagendr):agec + as.factor(riagendr):agec2, data = nhanes.red)


> svycollinear(mod=X, w=nhanes.red$wtmec2yr, Vcov=vcov(finmod), sc=TRUE, svyglm.obj=FALSE, rnd=3, fuzz=0.3)
   Cond.Index X.Intercept. as.factor.ridreth1.2 as.factor.ridreth1.3 as.factor.ridreth1.4
1           1            .                    .                    .                    .
2       1.332            .                    .                    .                    .
3       1.707            .                    .                    .                    .
4       1.729            .                    .                    .                    .
5       1.796            .                    .                    .                    .
6       2.199            .                    .                    .                    .
7       2.263            .                    .                    .                    .
8       2.298            .                    .                    .                    .
9       2.483            .                    .                    .                    .
10      2.733            .                    .                    .                    .
11      3.033            .                    .                    .                    .
12      3.206            .                    .                    .                    .
13      4.015            .                    .                    .                    .
14      4.217            .                    .                    .                    .
15      4.289            .                    .                    .                    .
```

16	4.327	.	.	.	.	
17	7.118	0.426	.	.	.	
18	8.984	.	.	.	0.433	
19	15.866	.	0.324	.	.	
20	24.225	0.566	0.631	0.573	0.532	

	as.factor.ridreth1.5	as.factor.riagendr.2	as.factor.marcat.2	as.factor.marcat.3	agec
1	.	.	.	.	.
2	.	.	.	.	.
3	.	.	.	.	.
4	.	.	.	.	.
5	.	.	.	.	.
6	.	.	.	.	.
7	.	.	.	.	.
8	.	.	.	.	.
9	.	.	.	.	.
10	.	.	.	.	.
11	.	.	.	.	.
12	.	.	.	.	0.877
13	.	.	.	.	.
14	.	.	0.645	.	.
15	.	0.314	.	.	.
16	.	.	.	0.342	.
17	.	0.469	.	.	.
18	0.324	.	.	.	.
19	.	.	.	.	.
20	0.645	.	.	.	.

	agec2	as.factor.ridreth1.2.agec	as.factor.ridreth1.3.agec	as.factor.ridreth1.4.agec
1	.	.	.	.
2	.	.	.	.
3	.	.	.	.
4	.	.	.	.
5	.	.	.	.
6	.	.	.	.
7	.	.	.	.
8	.	.	.	.
9	.	.	.	.
10	.	.	.	.
11	.	0.352	.	.
12	.	0.521	0.677	1
13	.	.	.	.
14	.	.	.	.
15	.	.	.	.
16	.	.	.	.
17	.	.	.	.
18	.	.	.	.
19	.	.	.	.
20	0.35	.	.	.

	as.factor.ridreth1.5.agec	as.factor.ridreth1.2.agec2	as.factor.ridreth1.3.agec2
1	.	.	.
2	.	.	.
3	.	.	.
4	.	.	.
5	.	0.368	.
6	.	0.305	0.338
7	.	.	.
8	.	.	.
9	.	.	.
10	.	.	.
11	.	.	.

12	0.615	.	.
13	.	.	.
14	.	.	.
15	.	.	.
16	.	.	.
17	.	.	.
18	.	.	.
19	.	.	.
20	.	.	0.386

	as.factor.ridreth1.4.agec2	as.factor.ridreth1.5.agec2	as.factor.riagendr.2.agec
1	.	.	.
2	.	.	.
3	.	.	.
4	.	.	.
5	.	.	.
6	0.308	.	.
7	.	.	.
8	.	.	.
9	.	.	0.625
10	.	.	.
11	.	.	.
12	.	.	.
13	.	.	.
14	.	.	.
15	.	.	.
16	.	.	.
17	.	.	.
18	.	.	.
19	.	.	.
20	0.375	0.368	.

	as.factor.riagendr.2.agec2
1	.
2	.
3	.
4	.
5	.
6	.
7	.
8	.
9	.
10	.
11	.
12	.
13	.
14	.
15	.
16	.
17	.
18	.
19	.
20	.

```
> # Investigate the variance inflation factors, largest are associated with age*age, not surprising
> svyvif(mobj=finmod, X=X[,-1], w=nhanes.red$wtmec2yr, stvar = "sdmvstra", clvar = "sdmvpsu")
$`Intercept adjusted`
                           svy.vif.m reg.vif.m     zeta     varrho.m zeta.x.varrho.m
as.factor(ridreth1)2        3.326801  3.635526  5281373 1.732658e-07       0.9150812
as.factor(ridreth1)3        4.510496  6.652603 11703317 5.793270e-08       0.6780048
as.factor(ridreth1)4        3.900630  4.552741  6108301 1.402624e-07       0.8567651
as.factor(ridreth1)5        2.964464  3.799036  6998364 1.115004e-07       0.7803202
as.factor(riagendr)2        1.779912  1.975032 13794492 6.533089e-08       0.9012064
as.factor(marcat)2          1.179270  1.171210 23613689 4.263977e-08       1.0068823
as.factor(marcat)3          1.020429  1.396216 14502291 5.039570e-08       0.7308531
agec                        6.682702 21.803199  6002225 5.106457e-08       0.3065010
agec2                      23.686454 23.293922  7653792 1.328559e-07       1.0168513
as.factor(ridreth1)2:agec   1.355413  2.364267  5697679 1.006184e-07       0.5732911
as.factor(ridreth1)3:agec   5.175231 15.291096  6372878 5.310746e-08       0.3384474
as.factor(ridreth1)4:agec   3.892812  3.295842  5037980 2.344448e-07       1.1811283
as.factor(ridreth1)5:agec   2.706873  2.559995  5352923 1.975321e-07       1.0573744
as.factor(ridreth1)2:agec2  7.178235  4.374053  8965329 1.830490e-07       1.6410946
as.factor(ridreth1)3:agec2 29.103191 22.673034  9051476 1.418115e-07       1.2836037
as.factor(ridreth1)4:agec2  7.006965  5.916054  5386111 2.198986e-07       1.1843984
as.factor(ridreth1)5:agec2  5.869827  4.647206  5039525 2.506362e-07       1.2630873
as.factor(riagendr)2:agec   1.857950  2.378967 11624875 6.718267e-08       0.7809901
as.factor(riagendr)2:agec2  4.924833  3.565446 12607801 1.095566e-07       1.3812673
                              Rsq.m
as.factor(ridreth1)2       0.7249366
as.factor(ridreth1)3       0.8496829
as.factor(ridreth1)4       0.7803521
as.factor(ridreth1)5       0.7367753
as.factor(riagendr)2       0.4936791
as.factor(marcat)2         0.1461818
as.factor(marcat)3         0.2837784
agec                       0.9541352
agec2                      0.9570703
as.factor(ridreth1)2:agec  0.5770360
as.factor(ridreth1)3:agec  0.9346025
as.factor(ridreth1)4:agec  0.6965874
as.factor(ridreth1)5:agec  0.6093743
as.factor(ridreth1)2:agec2 0.7713791
as.factor(ridreth1)3:agec2 0.9558947
as.factor(ridreth1)4:agec2 0.8309684
as.factor(ridreth1)5:agec2 0.7848170
as.factor(riagendr)2:agec  0.5796495
as.factor(riagendr)2:agec2 0.7195301

$`No intercept`
                             svy.vif   reg.vif     zeta       varrho zeta.x.varrho       Rsq
as.factor(ridreth1)2        4.189635  3.886793  5281373 2.040976e-07     1.0779157 0.7427185
as.factor(ridreth1)3        8.933934 19.936135 11703317 3.829066e-08     0.4481277 0.9498398
as.factor(ridreth1)4       10.073420  5.144949  6108301 3.205350e-07     1.9579240 0.8056346
as.factor(ridreth1)5        4.831965  4.117288  6998364 1.676934e-07     1.1735796 0.7571217
as.factor(riagendr)2        3.189268  4.033431 13794492 5.732060e-08     0.7907085 0.7520721
as.factor(marcat)2          1.511311  1.439206 23613689 4.446997e-08     1.0501000 0.3051725
as.factor(marcat)3          1.657764  1.745576 14502291 6.548584e-08     0.9496946 0.4271233
agec                        6.672543 21.897485  6002225 5.076740e-08     0.3047173 0.9543327
agec2                      50.503972 44.467672  7653792 1.483900e-07     1.1357458 0.9775118
as.factor(ridreth1)2:agec   1.350741  2.370101  5697679 1.000247e-07     0.5699088 0.5780770
as.factor(ridreth1)3:agec   5.310793 15.663268  6372878 5.320364e-08     0.3390603 0.9361564
as.factor(ridreth1)4:agec   3.985194  3.299285  5037980 2.397580e-07     1.2078961 0.6969040
as.factor(ridreth1)5:agec   2.715103  2.562989  5352923 1.979013e-07     1.0593502 0.6098306
```

```
as.factor(ridreth1)2:agec2  8.173427  4.524598  8965329 2.014921e-07     1.8064427 0.7789859
as.factor(ridreth1)3:agec2 29.636740 33.096026  9051476 9.893163e-08     0.8954773 0.9697849
as.factor(ridreth1)4:agec2 12.171412  6.263483  5386111 3.607861e-07     1.9432339 0.8403444
as.factor(ridreth1)5:agec2  9.267230  4.830739  5039525 3.806683e-07     1.9183875 0.7929923
as.factor(riagendr)2:agec   1.847998  2.392992 11624875 6.643117e-08     0.7722540 0.5821131
as.factor(riagendr)2:agec2  6.010733  4.705570 12607801 1.013155e-07     1.2773656 0.7874859
```

```
> # Compute standardized residuals
> st.resids <- svystdres(mobj = finmod, stvar = "sdmvstra", clvar = "sdmvpsu", doplot = TRUE)
> # see automatically generated dot plot

> st.resids$stdresids[st.resids$stdresids < -4]
     2735      3732      3138      1158      5815
-4.186686 -5.030542 -4.179745 -4.036181 -4.826700
> # examine residuals < -4
> #     2735      3732      3138      1158      5815
> # -4.186686 -5.030542 -4.179745 -4.036181 -4.826700
```

```
> # Re-run  Final Model including interactions of race X age, gender X age, and main effects of marital status
> ex75_final_2 <- svyglm(bpxdi1_1 ~ marcatc + agec*factor(racec) + agecsq*factor(racec) + factor(genderc)*agec +
factor(genderc)*agecsq,
+   subnhanes, family=gaussian)
>
> # Check coefficients
> (ex75_final_2$coefficients)
                       (Intercept)          marcatcPreviously Married
                      74.9279513829                       0.2905908093
              marcatcNever Married                               agec
                       1.1076854953                       0.0281912152
       factor(racec)Other Hispanic                  factor(racec)White
                       0.4060606458                       1.6140818145
                 factor(racec)Black                  factor(racec)Other
                       3.5198660800                       1.4164313499
                             agecsq                    factor(genderc)F
                      -0.0138640853                      -3.1027737781
  agec:factor(racec)Other Hispanic             agec:factor(racec)White
                       0.0605771656                       0.0076903566
           agec:factor(racec)Black             agec:factor(racec)Other
                       0.0646497179                       0.0390183239
factor(racec)Other Hispanic:agecsq          factor(racec)White:agecsq
                      -0.0003223620                       0.0015387958
         factor(racec)Black:agecsq          factor(racec)Other:agecsq
                      -0.0042948664                      -0.0003766444
             agec:factor(genderc)F            agecsq:factor(genderc)F
                       0.0484726070                       0.0030104518
>
> # Obtain summary of model
> summary(ex75_final_2, df.resid=degf(subnhanes))

Call:
svyglm(formula = bpxdi1_1 ~ marcatc + agec * factor(racec) +
    agecsq * factor(racec) + factor(genderc) * agec + factor(genderc) *
    agecsq, design = subnhanes, family = gaussian)
```

```
Survey design:
subset(nhanessvy2, age >= 18)

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                        74.9279514  0.9327484  80.330  < 2e-16 ***
marcatcPreviously Married           0.2905908  0.6398622   0.454 0.655469
marcatcNever Married                1.1076855  0.7757962   1.428 0.171461
agec                                0.0281912  0.0397675   0.709 0.487995
factor(racec)Other Hispanic         0.4060606  0.9797099   0.414 0.683713
factor(racec)White                  1.6140818  0.9440749   1.710 0.105510
factor(racec)Black                  3.5198661  0.9909619   3.552 0.002452 **
factor(racec)Other                  1.4164313  0.8617961   1.644 0.118628
agecsq                             -0.0138641  0.0023774  -5.832    2e-05 ***
factor(genderc)F                   -3.1027738  0.7561240  -4.104 0.000741 ***
agec:factor(racec)Other Hispanic    0.0605772  0.0519876   1.165 0.260019
agec:factor(racec)White             0.0076904  0.0564603   0.136 0.893257
agec:factor(racec)Black             0.0646497  0.0391346   1.652 0.116887
agec:factor(racec)Other             0.0390183  0.0620544   0.629 0.537853
factor(racec)Other Hispanic:agecsq -0.0003224  0.0036160  -0.089 0.930005
factor(racec)White:agecsq           0.0015388  0.0019529   0.788 0.441566
factor(racec)Black:agecsq          -0.0042949  0.0023828  -1.802 0.089235 .
factor(racec)Other:agecsq          -0.0003766  0.0024760  -0.152 0.880885
agec:factor(genderc)F               0.0484726  0.0259984   1.864 0.079620 .
agecsq:factor(genderc)F             0.0030105  0.0016991   1.772 0.094355 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero or negative residual df; p-values not defined

(Dispersion parameter for gaussian family taken to be 123.3161)

Number of Fisher Scoring iterations: 2

>
> # Calculate CI as follows
> #lowerci
> coef(ex75_final_2) - (1.96*SE(ex75_final_2))
                       (Intercept)          marcatcPreviously Married
                     73.0997646160                       -0.9635391803
              marcatcNever Married                                agec
                     -0.4128751141                       -0.0497530847
       factor(racec)Other Hispanic                 factor(racec)White
                     -1.5141706938                       -0.2363050453
                factor(racec)Black                 factor(racec)Other
                      1.5775806627                       -0.2726890581
                             agecsq                   factor(genderc)F
                     -0.0185237765                       -4.5847768970
  agec:factor(racec)Other Hispanic            agec:factor(racec)White
                     -0.0413185847                       -0.1029717557
           agec:factor(racec)Black            agec:factor(racec)Other
                     -0.0120541846                       -0.0826083357
factor(racec)Other Hispanic:agecsq          factor(racec)White:agecsq
                     -0.0074097000                       -0.0022888331
         factor(racec)Black:agecsq          factor(racec)Other:agecsq
                     -0.0089651861                       -0.0052296027
             agec:factor(genderc)F            agecsq:factor(genderc)F
                     -0.0024841923                       -0.0003198212
>
```

```
> #upperci
> coef(ex75_final_2) + (1.96*SE(ex75_final_2))
                       (Intercept)          marcatcPreviously Married
                     76.7561381498                       1.5447207990
              marcatcNever Married                               agec
                      2.6282461046                       0.1061355151
       factor(racec)Other Hispanic                 factor(racec)White
                      2.3262919855                       3.4644686743
                factor(racec)Black                 factor(racec)Other
                      5.4621514972                       3.1055517579
                             agecsq                   factor(genderc)F
                     -0.0092043942                      -1.6207706591
  agec:factor(racec)Other Hispanic            agec:factor(racec)White
                      0.1624729160                       0.1183524690
           agec:factor(racec)Black            agec:factor(racec)Other
                      0.1413536204                       0.1606449834
factor(racec)Other Hispanic:agecsq          factor(racec)White:agecsq
                      0.0067649761                       0.0053664246
         factor(racec)Black:agecsq          factor(racec)Other:agecsq
                      0.0003754534                       0.0044763139
             agec:factor(genderc)F            agecsq:factor(genderc)F
                      0.0994294062                       0.0063407249
```

```
> # 7.5.9 Q approach for weighting, see Pfefferman for details, table 7.6 column 2
> # Use complete case data frame called nhanes.red from diagnostics analysis above
> # Step 1 linear model with weight regressed on race, gender marcat and age / gender interactions
> # In complete case adult data set >= 18 years of age

> q_wgt <- lm(wtmec2yr ~ factor(marcat) + agec*factor(ridreth1) + agec2*factor(ridreth1) +
+   factor(riagendr)*agec + factor(riagendr)*agec2 , nhanes.red)
> summary(q_wgt)

Call:
lm(formula = wtmec2yr ~ factor(marcat) + agec * factor(ridreth1) +
    agec2 * factor(ridreth1) + factor(riagendr) * agec + factor(riagendr) *
    agec2, data = nhanes.red)

Residuals:
   Min     1Q Median     3Q    Max
-79115  -7934  -1632   6748 156982

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              3.353e+04  1.965e+03  17.061  < 2e-16 ***
factor(marcat)2         -8.498e+03  1.117e+03  -7.608 3.33e-14 ***
factor(marcat)3         -2.680e+02  1.205e+03  -0.222  0.82396
agec                    -6.852e+02  8.954e+01  -7.652 2.37e-14 ***
factor(ridreth1)2       -4.209e+03  2.681e+03  -1.570  0.11647
factor(ridreth1)3        6.206e+04  2.134e+03  29.084  < 2e-16 ***
factor(ridreth1)4       -1.203e+04  2.229e+03  -5.396 7.13e-08 ***
factor(ridreth1)5       -1.357e+04  2.374e+03  -5.717 1.15e-08 ***
agec2                   -4.338e+00  5.524e+00  -0.785  0.43226
factor(riagendr)2        1.259e+03  1.196e+03   1.053  0.29262
agec:factor(ridreth1)2   5.812e+01  1.175e+02   0.495  0.62094
agec:factor(ridreth1)3   8.414e+02  9.704e+01   8.671  < 2e-16 ***
agec:factor(ridreth1)4   4.976e+02  9.897e+01   5.028 5.14e-07 ***
agec:factor(ridreth1)5   6.517e+02  1.061e+02   6.140 8.91e-10 ***
factor(ridreth1)2:agec2  3.426e+00  7.203e+00   0.476  0.63434
factor(ridreth1)3:agec2 -4.800e+01  5.825e+00  -8.240  < 2e-16 ***
factor(ridreth1)4:agec2 -3.558e+00  6.062e+00  -0.587  0.55720
factor(ridreth1)5:agec2 -8.281e-01  6.506e+00  -0.127  0.89873
agec:factor(riagendr)2   9.405e+01  5.111e+01   1.840  0.06581 .
agec2:factor(riagendr)2  9.025e+00  2.811e+00   3.211  0.00133 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28770 on 4825 degrees of freedom
Multiple R-squared:  0.5097,    Adjusted R-squared:  0.5078
F-statistic:   264 on 19 and 4825 DF,  p-value: < 2.2e-16


> w_hat <- predict(q_wgt)
> summary(w_hat)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -2927   19988   25563   42411   70708   97203


> # Set values < 0 to 4809 (1 percentile value)
> w_hat <- ifelse (w_hat < 0 , 4809,w_hat)
> summary(w_hat)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  509.2 19988.0 25563.0 42446.3 70708.4 97203.0
```

```
> # Create revised weight
> nhanes.red$q_wtmec2yr = (nhanes.red$wtmec2yr / w_hat)

> # Design object for analysis
> nhanessvyq <- svydesign(strata=~sdmvstra, id=~sdmvpsu, weights=~q_wtmec2yr, data=nhanes.red, nest=T)


> # Final model with revised Q weight, table 7.6
> ex75_finalq <- svyglm(bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) + as.factor(marcat) +
+   agec + agec2 + as.factor(ridreth1):agec + as.factor(ridreth1):agec2 +
+    as.factor(riagendr):agec + as.factor(riagendr):agec2, design = nhanessvyq)
> summary(ex75_finalq, df=NULL,df.resid=Inf )

Call:
svyglm(formula = bpxdi1.1 ~ as.factor(ridreth1) + as.factor(riagendr) +
    as.factor(marcat) + agec + agec2 + as.factor(ridreth1):agec +
    as.factor(ridreth1):agec2 + as.factor(riagendr):agec + as.factor(riagendr):agec2,
    design = nhanessvyq)

Survey design:
svydesign(strata = ~sdmvstra, id = ~sdmvpsu, weights = ~q_wtmec2yr,
    data = nhanes.red, nest = T)

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                75.3946699  0.9674261  77.933  < 2e-16 ***
as.factor(ridreth1)2        0.7803662  0.9815935   0.795  0.42661
as.factor(ridreth1)3        1.4423700  1.0601621   1.361  0.17367
as.factor(ridreth1)4        3.4278898  1.0558833   3.246  0.00117 **
as.factor(ridreth1)5        1.2152539  0.9921961   1.225  0.22065
as.factor(riagendr)2       -3.4327530  0.6387017  -5.375 7.68e-08 ***
as.factor(marcat)2         -0.2029976  0.5186602  -0.391  0.69551
as.factor(marcat)3          0.5518717  0.5901093   0.935  0.34968
agec                        0.0438633  0.0431467   1.017  0.30934
agec2                      -0.0138047  0.0027797  -4.966 6.82e-07 ***
as.factor(ridreth1)2:agec   0.0172068  0.0642658   0.268  0.78890
as.factor(ridreth1)3:agec  -0.0042595  0.0595282  -0.072  0.94296
as.factor(ridreth1)4:agec   0.0704203  0.0391069   1.801  0.07175 .
as.factor(ridreth1)5:agec   0.0371382  0.0595042   0.624  0.53254
as.factor(ridreth1)2:agec2 -0.0036177  0.0049365  -0.733  0.46366
as.factor(ridreth1)3:agec2  0.0017947  0.0022662   0.792  0.42840
as.factor(ridreth1)4:agec2 -0.0032639  0.0030665  -1.064  0.28716
as.factor(ridreth1)5:agec2  0.0006603  0.0028969   0.228  0.81969
as.factor(riagendr)2:agec   0.0309103  0.0298147   1.037  0.29985
as.factor(riagendr)2:agec2  0.0028465  0.0018944   1.503  0.13295
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero or negative residual df; p-values not defined

(Dispersion parameter for gaussian family taken to be 122.9104)

Number of Fisher Scoring iterations: 2
```

```
> # Alternative Bayesian Approach

> # load/install needed packages if not already in place
> library(devtools)
Loading required package: usethis
Warning message:
package ‘devtools’ was built under R version 4.4.3
> #install_github("RyanHornby/csSampling") # lots of packages needed, will take a second
> library(csSampling)
Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
> library(rstan)
Loading required package: StanHeaders

rstan version 2.32.6 (Stan version 2.32.2)

For execution on a local, multicore CPU with excess RAM we recommend calling
options(mc.cores = parallel::detectCores()).
To avoid recompilation of unchanged Stan programs, we recommend calling
rstan_options(auto_write = TRUE)
For within-chain threading using `reduce_sum()` or `map_rect()` Stan functions,
change `threads_per_chain` option:
rstan_options(threads_per_chain = 1)

Do not specify '-march=native' in 'LOCAL_CPPFLAGS' or a Makevars file
> library(brms)
Loading required package: Rcpp
Loading 'brms' package (version 2.21.0). Useful instructions
can be found by typing help('brms'). A more detailed introduction
to the package is available through vignette('brms_overview').

Attaching package: ‘brms’

The following object is masked from ‘package:rstan’:

    loo

The following object is masked from ‘package:survival’:

    kidney

The following object is masked from ‘package:stats’:

    ar

> library(survey)
> rstan_options(auto_write = TRUE)
>
> # Need to normalize NHANES weights to match what is done for Stan modeling
> nhanes.red$wtsc <- nhanes.red$wtmec2yr / mean(nhanes.red$wtmec2yr)
>
>
> # Survey design object
> nhanes.des <- svydesign(id = ~sdmvpsu, strata = ~sdmvstra, weights = ~wtsc, nest = T, data = nhanes.red)
>
>
```

```
> # Bayesian approach, flat prior
> set.seed(41279)
> model_formula <- formula("bpxdi1.1|weights(wtsc) ~ as.factor(ridreth1) + as.factor(riagendr) +
as.factor(marcat) + agec + agec2 + as.factor(ridreth1):agec + as.factor(ridreth1):agec2 +
as.factor(riagendr):agec + as.factor(riagendr):agec2")
> mod.brms <- cs_sampling_brms(svydes = nhanes.des, brmsmod = brmsformula(model_formula), data = nhanes.red,
family = gaussian())
[1] "compiling stan model"
[1] "stan fitting"

SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).
Chain 1:
Chain 1: Gradient evaluation took 0.001192 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 11.92 seconds.
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 1: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 1: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 1: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 1: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 1: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 1: Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 1: Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 1: Iteration: 2000 / 2000 [100%]  (Sampling)
Chain 1:
Chain 1:  Elapsed Time: 99.235 seconds (Warm-up)
Chain 1:                51.581 seconds (Sampling)
Chain 1:                150.816 seconds (Total)
Chain 1:
[1] "gradient evaluation"
Warning message:
In mrbweights(design$cluster, design$strata, design$fpc, ...) :
  Design is sampled with replacement: only first stage used
```

```
> # Print results
> mod.brms$stan_fit
Inference for Stan model: anon_model.
1 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=1000.

                mean se_mean   sd      2.5%       25%       50%       75%     97.5% n_eff
b[1]            0.40    0.05 1.19     -1.94     -0.37      0.44      1.18      2.73   534
b[2]            1.57    0.04 0.85     -0.11      1.01      1.59      2.14      3.31   391
b[3]            3.49    0.05 1.05      1.51      2.79      3.49      4.18      5.61   451
b[4]            1.34    0.06 1.16     -0.97      0.56      1.33      2.14      3.47   427
b[5]           -3.09    0.02 0.44     -3.91     -3.38     -3.08     -2.80     -2.19   841
b[6]            0.27    0.01 0.42     -0.54     -0.03      0.27      0.57      1.11   970
b[7]            1.10    0.01 0.44      0.28      0.79      1.11      1.39      1.95  1022
b[8]            0.03    0.00 0.04     -0.04      0.00      0.03      0.06      0.11   338
b[9]           -0.01    0.00 0.00     -0.02     -0.02     -0.01     -0.01     -0.01   369
b[10]           0.06    0.00 0.05     -0.05      0.02      0.06      0.09      0.16   443
b[11]           0.00    0.00 0.04     -0.08     -0.02      0.01      0.03      0.09   350
b[12]           0.06    0.00 0.05     -0.04      0.03      0.06      0.09      0.15   391
b[13]           0.04    0.00 0.05     -0.06      0.00      0.04      0.07      0.13   402
b[14]           0.00    0.00 0.00     -0.01      0.00      0.00      0.00      0.01   457
b[15]           0.00    0.00 0.00      0.00      0.00      0.00      0.00      0.01   406
b[16]           0.00    0.00 0.00     -0.01     -0.01      0.00      0.00      0.00   475
b[17]           0.00    0.00 0.00     -0.01      0.00      0.00      0.00      0.01   417
b[18]           0.05    0.00 0.02      0.01      0.04      0.05      0.06      0.09   722
b[19]           0.00    0.00 0.00      0.00      0.00      0.00      0.00      0.01   834
Intercept      71.56    0.01 0.19     71.19     71.42     71.55     71.70     71.95   897
sigma          10.75    0.00 0.11     10.52     10.68     10.75     10.83     10.97   661
lprior         -6.74    0.00 0.01     -6.76     -6.75     -6.74     -6.74     -6.73   667
b_Intercept    74.96    0.04 0.84     73.21     74.42     74.92     75.49     76.72   387
lp__        -18384.34    0.17 3.18 -18391.80 -18386.20 -18384.01 -18382.08 -18378.88   351
            Rhat
b[1]           1
b[2]           1
b[3]           1
b[4]           1
b[5]           1
b[6]           1
b[7]           1
b[8]           1
b[9]           1
b[10]          1
b[11]          1
b[12]          1
b[13]          1
b[14]          1
b[15]          1
b[16]          1
b[17]          1
b[18]          1
b[19]          1
Intercept      1
sigma          1
lprior         1
b_Intercept    1
lp__           1
```

Samples were drawn using NUTS(diag_e) at Sun Apr 13 12:52:11 2025.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).