Svyset analysis

Chaochen Wang

25 July 2018

Contents

```
statapath <- "/usr/local/stata15/stata" # <- Ubuntu path to stata
#statapath <- "D:/software/STATA/Stata-64.exe" # <- Windows STATA path modify if needed
knitr::opts_chunk$set(engine.path = list(
  stata = statapath
))
library(naniar)
library(epiDisplay)
library(readr)
library(plyr)
library(dplyr)
library(tidyverse)
library(haven)
library(survey)
 # install packages if needed
```

Read the datasets and prepare datasets

CW2CB2 data set (2by2 multilevel classes)

```
CW2CB2 <- read_table2("~/Documents/LSHTMproject/results/50NDNS_CW2CB2.txt",
                        col_names = FALSE)# change the path to your own path


names(CW2CB2) <- c("H0", "H1", "H2", "H3", "H4", "H5", "H6", "H7",
                   "H8", "H9", "H10", "H11", "H12", "H13", "H14",
                   "H15", "H16", "H17", "H18", "H19", "H20", "H21",
                   "H22", "H23", "ID_DAY", "AGE", "SEX", "CPROB1",
                   "CPROB2", "CPROB3", "CPROB4",
```

```r
                 # "CPROB5",
                 # "CPROB6",
                 "CB", "CW", "MLCJOINT", "ID")
CW2CB2_reg <- CW2CB2[!duplicated(CW2CB2$ID), ]

CW2CB2_reg <- CW2CB2_reg %>%
  select(ID, AGE, SEX, CB) # extract only the CB variable (Between individual classes == 1 or 2)

tab1(CW2CB2_reg$CB, graph = FALSE)
```

```
## CW2CB2_reg$CB :
##         Frequency Percent Cum. percent
## 1            4147    67.4         67.4
## 2            2008    32.6        100.0
##   Total      6155   100.0        100.0
```

CW3CB2 data set (3by2 multilevel classes)

```r
CW3CB2 <- read_table2("~/Documents/LSHTMproject/results/50NDNS_CW3CB2.txt",
                    col_names = FALSE)# change the path to your own path

names(CW3CB2) <- c("H0", "H1", "H2", "H3", "H4", "H5", "H6", "H7",
                "H8", "H9", "H10", "H11", "H12", "H13", "H14",
                "H15", "H16", "H17", "H18", "H19", "H20", "H21",
                "H22", "H23", "ID_DAY", "AGE", "SEX", "CPROB1",
                "CPROB2", "CPROB3", "CPROB4", "CPROB5", "CPROB6",
                "CB", "CW", "MLCJOINT", "ID")
CW3CB2_reg <- CW3CB2[!duplicated(CW3CB2$ID), ]

CW3CB2_reg <- CW3CB2_reg %>%
  select(ID, AGE, SEX, CB) # extract only the CB variable (Between individual classes == 1 or 2)

tab1(CW3CB2_reg$CB, graph = FALSE)
```

```
## CW3CB2_reg$CB :
##         Frequency Percent Cum. percent
## 1            3743    60.8         60.8
## 2            2412    39.2        100.0
##   Total      6155   100.0        100.0
```

LCGA data set (2 classes)

```r
CARB_50_LGCA_2CLASS <- read_table2("~/Documents/LSHTMproject/results/LCGA/CARB_50_LGCA_2CLASS.txt",
    col_names = FALSE)# change the path to your own path

names(CARB_50_LGCA_2CLASS) <- c("H0_X",  "H1_X", "H2_X", "H3_X", "H4_X",  "H5_X", "H6_X",  "H7_X",
  "H8_X",  "H9_X", "H10_X",  "H11_X", "H12_X",  "H13_X", "H14_X",  "H15_X", "H16_X",  "H17_X",
  "H18_X",  "H19_X", "H20_X",  "H21_X", "H22_X",  "H23_X", "H0_Y",  "H1_Y", "H2_Y",  "H3_Y",
  "H4_Y",  "H5_Y", "H6_Y",  "H7_Y", "H8_Y",  "H9_Y", "H10_Y",  "H11_Y", "H12_Y",  "H13_Y",
  "H14_Y",  "H15_Y", "H16_Y",  "H17_Y", "H18_Y",  "H19_Y", "H20_Y",  "H21_Y",
  "H22_Y",  "H23_Y", "H0_X_X",  "H1_X_X", "H2_X_X",  "H3_X_X", "H4_X_X",  "H5_X_X",
```

```
  "H6_X_X",   "H7_X_X",  "H8_X_X",   "H9_X_X",  "H10_X_X",   "H11_X_X",  "H12_X_X",   "H13_X_X",
  "H14_X_X",   "H15_X_X", "H16_X_X",  "H17_X_X", "H18_X_X",   "H19_X_X", "H20_X_X",   "H21_X_X",
  "H22_X_X",   "H23_X_X", "H0_Y_Y",   "H1_Y_Y",  "H2_Y_Y",    "H3_Y_Y",  "H4_Y_Y",    "H5_Y_Y",
  "H6_Y_Y",   "H7_Y_Y",  "H8_Y_Y",   "H9_Y_Y",  "H10_Y_Y",   "H11_Y_Y",  "H12_Y_Y",   "H13_Y_Y",
  "H14_Y_Y",   "H15_Y_Y", "H16_Y_Y",  "H17_Y_Y", "H18_Y_Y",   "H19_Y_Y", "H20_Y_Y",   "H21_Y_Y",
  "H22_Y_Y",   "H23_Y_Y", "ID", "CPROB1",   "CPROB2",
  #"CPROB3",   #"CPROB4",
  "C")

CARB_50_LGCA_2CLASS[CARB_50_LGCA_2CLASS == "*"] <- NA

LCGA_2class <- CARB_50_LGCA_2CLASS %>%
  select(ID, C) # extract only the C variable (classes == 1 or 2)

tab1(LCGA_2class$C, graph = FALSE)
```

```
## LCGA_2class$C :
##         Frequency Percent Cum. percent
## 1            4283    69.6         69.6
## 2            1872    30.4        100.0
##   Total      6155   100.0        100.0
```

LCGA data set (3 classes)

```
CARB_50_LGCA_3CLASS <- read_table2("~/Documents/LSHTMproject/results/LCGA/CARB_50_LGCA_3CLASS.DAT",
    col_names = FALSE)# change the path to your own path

names(CARB_50_LGCA_3CLASS) <- c("H0_X",  "H1_X", "H2_X", "H3_X", "H4_X", "H5_X", "H6_X",  "H7_X",
  "H8_X",  "H9_X",  "H10_X",  "H11_X", "H12_X",  "H13_X", "H14_X",  "H15_X", "H16_X",  "H17_X",
  "H18_X",  "H19_X", "H20_X",  "H21_X", "H22_X",  "H23_X", "H0_Y",  "H1_Y", "H2_Y",  "H3_Y",
  "H4_Y",  "H5_Y", "H6_Y",  "H7_Y", "H8_Y",  "H9_Y", "H10_Y",  "H11_Y", "H12_Y",  "H13_Y",
  "H14_Y",  "H15_Y", "H16_Y",  "H17_Y", "H18_Y",  "H19_Y", "H20_Y",  "H21_Y",
  "H22_Y",  "H23_Y", "H0_X_X",  "H1_X_X", "H2_X_X",  "H3_X_X", "H4_X_X",  "H5_X_X",
  "H6_X_X",  "H7_X_X", "H8_X_X",  "H9_X_X", "H10_X_X",  "H11_X_X", "H12_X_X",   "H13_X_X",
  "H14_X_X",   "H15_X_X", "H16_X_X",  "H17_X_X", "H18_X_X",   "H19_X_X", "H20_X_X",   "H21_X_X",
  "H22_X_X",   "H23_X_X", "H0_Y_Y",   "H1_Y_Y",  "H2_Y_Y",    "H3_Y_Y",  "H4_Y_Y",    "H5_Y_Y",
  "H6_Y_Y",   "H7_Y_Y",  "H8_Y_Y",   "H9_Y_Y",  "H10_Y_Y",   "H11_Y_Y",  "H12_Y_Y",   "H13_Y_Y",
  "H14_Y_Y",   "H15_Y_Y", "H16_Y_Y",  "H17_Y_Y", "H18_Y_Y",   "H19_Y_Y", "H20_Y_Y",   "H21_Y_Y",
  "H22_Y_Y",   "H23_Y_Y", "ID", "CPROB1",   "CPROB2",
  "CPROB3",   "C")

CARB_50_LGCA_3CLASS[CARB_50_LGCA_3CLASS == "*"] <- NA

LCGA_3class <- CARB_50_LGCA_3CLASS %>%
  select(ID, C) # extract only the C variable (classes == 1 or 2)

tab1(LCGA_3class$C, graph = FALSE)
```

```
## LCGA_3class$C :
##         Frequency Percent Cum. percent
## 1            1783    29.0         29.0
## 2             376     6.1         35.1
## 3            3996    64.9        100.0
```

```
##    Total       6155    100.0          100.0
```

Individual data (BMI, WC, blood pressure, A1C etc.)

```r
# change the following path according to your own data folders

blood78 <- read_dta("~/Downloads/UKDA-6533-stata11_se/stata11_se/ndns_rp_yr7-8a_indiv.dta")
blood56 <- read_dta("~/Downloads/UKDA-6533-stata11_se/stata11_se/ndns_rp_yr5-6a_indiv.dta")
blood14 <- read_dta("~/Downloads/UKDA-6533-stata11_se/stata11_se/ndns_rp_yr1-4a_indiv_uk.dta")

names(blood78)[names(blood78)=="seriali"] <- "ID"
names(blood56)[names(blood56)=="seriali"] <- "ID"
names(blood14)[names(blood14)=="seriali"] <- "ID"

BMI78 <- blood78 %>%
  select(ID, Sex, age, bmival, wstval, Diabetes, bpmedc2, bpmedd2, hyper140_2, hibp140_2,
         Glucose, A1C, cigsta3, dnoft3, dnnow, wti_Y78, wtn_Y78, wtb_Y78, cluster1, cluster2, cluster3,
         cluster4, cluster5, area, gor) %>%
  rename(wti = wti_Y78, wtn = wtn_Y78, wtb = wtb_Y78, drink = dnoft3) %>%
  mutate(Years = "7-8") %>%
  replace_with_na(replace = list(bmival = -1,
                                 wstval = -1,
                                 bpmedd2 = -1,
                                 bpmedc2 = -1,
                                 hyper140_2 = -7,
                                 # hyper140_2 = -1,
                                 hibp140_2 = -7,
                                 # hibp140_2 = -1,
                                 Glucose = -1,
                                 A1C  = -1,
                                 dnnow = -1,
                                 drink = -1,
                                 cigsta3 = -1)) %>%
  replace_with_na(replace = list(hyper140_2 = -1, hibp140_2 = -1,
                                 drink = -8)) %>%
  replace_with_na(replace = list(drink = -9))




BMI56 <- blood56 %>%
  select(ID, Sex, age, area, bmival, wstval, Diabetes, bpmedc2, bpmedd2, hyper140_2, hibp140_2,
         Glucose, A1C, cigsta3, dnoft3, dnnow, wti_Y56, wtn_Y56, wtb_Y56, cluster1, cluster2, cluster3,
         cluster4, cluster5, area, gor) %>%
  mutate(Years = "5-6") %>%
  rename(wti = wti_Y56, wtn = wtn_Y56, wtb = wtb_Y56, drink = dnoft3) %>%
  replace_with_na(replace = list(bmival = -1,
                                 wstval = -1,
                                 bpmedd2 = -1,
                                 bpmedc2 = -1,
                                 hyper140_2 = -7,
                                 # hyper140_2 = -1,
                                 hibp140_2 = -7,
```

```r
                                            # hibp140_2 = -1,
                                            Glucose = -1,
                                            A1C   = -1,
                                            dnnow = -1,
                                            drink = -1,
                                            cigsta3 = -1)) %>%
  replace_with_na(replace = list(hyper140_2 = -1, hibp140_2 = -1,
                                 drink = -8)) %>%
  replace_with_na(replace = list(drink = -9))


BMI14 <- blood14 %>%
  select(ID, Sex, age, bmival, wstval, Diabetes, bpmedc, bpmedd, hyper140, hibp140,
         Glucose, A1C, cigsta3, dnoft3, dnnow, wti_CY1234, wtn_CY1234, wtb_CY1234, cluster, area, gor)
  rename(hyper140_2 = hyper140, hibp140_2 = hibp140, bpmedd2 = bpmedd,
         bpmedc2 = bpmedc, cluster1 = cluster,
         wti = wti_CY1234, wtn = wtn_CY1234, wtb = wtb_CY1234, drink = dnoft3) %>%
  mutate(cluster2 = NA, cluster3 = NA, cluster4 = NA, cluster5 = NA, Years = "1-4") %>%
  replace_with_na(replace = list(bmival = -1,
                                 wstval = -1,
                                 bpmedd2 = -1,
                                 bpmedc2 = -1,
                                 hyper140_2 = -7,
                                 # hyper140_2 = -1,
                                 hibp140_2 = -7,
                                 # hibp140_2 = -1,
                                 Glucose = -1,
                                 A1C   = -1,
                                 dnnow = -1,
                                 drink = -1,
                                 cigsta3 = -1)) %>%
  replace_with_na(replace = list(hyper140_2 = -1, hibp140_2 = -1,
                                 drink = -8)) %>%
  replace_with_na(replace = list(drink = -9))



BMI <- bind_rows(BMI14, BMI56, BMI78)

CW2CB2_regss <- CW2CB2_reg %>%
  left_join(BMI, by = "ID") ## dataset for 2by2 multilevel latent classes

CW3CB2_regss <- CW3CB2_reg %>%
  left_join(BMI, by = "ID") ## dataset for 3by2 multilevel latent classes

LCGA_2class <- LCGA_2class %>%
  left_join(BMI, by = "ID") ## dataset for 2 classes LCGA

LCGA_3class <- LCGA_3class %>%
  left_join(BMI, by = "ID") ## dataset for 3 classes LCGA

rm(blood14, blood56, blood78, BMI14, BMI56, BMI78, #BMI,
   CW2CB2, CW2CB2_reg, CW3CB2, CW3CB2_reg, CARB_50_LGCA_2CLASS, CARB_50_LGCA_3CLASS)
```

Rescale the weighting values

```
# individual weighting
a <- sum(CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wti)
b <- sum(CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wti)
c <- sum(CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wti)

CW2CB2_regss$wti1to8 <- CW2CB2_regss$wti

CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wti1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wti*(a-
CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wti1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wti*(a-
CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wti1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wti*(a-
mean(CW2CB2_regss$wti1to8)
```

```
## [1] 1.209817
```

```
CW2CB2_regss$wti1to8 <- CW2CB2_regss$wti1to8/1.209816814
summ(CW2CB2_regss$wti1to8, graph = FALSE)
```

```
##  obs. mean   median  s.d.   min.   max.
##  6155 1       0.892   0.897  0      5.893
```

```
sum(CW2CB2_regss$wti1to8, graph = FALSE) #Check if the weighting sum up to the sample size we have
```

```
## [1] 6155
```

```
# Nurse weights

a <- sum(CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wtn)
b <- sum(CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wtn)
c <- sum(CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wtn)

CW2CB2_regss$wtn1to8 <- CW2CB2_regss$wtn

CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wtn1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wtn*(a-
CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wtn1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wtn*(a-
CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wtn1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wtn*(a-
mean(CW2CB2_regss$wtn1to8)
```

```
## [1] 0.9070036
```

```
CW2CB2_regss$wtn1to8 <- CW2CB2_regss$wtn1to8/0.907003577
summ(CW2CB2_regss$wtn1to8, graph = FALSE)
```

```
##  obs. mean   median  s.d.   min.   max.
##  6155 1       0.588   1.203  0      8.516
```

```
sum(CW2CB2_regss$wtn1to8, graph = FALSE) #Check if the weighting sum up to the sample size we have
```

```
## [1] 6155
```

```
# Blood weights
a <- sum(CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wtb)
b <- sum(CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wtb)
c <- sum(CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wtb)

CW2CB2_regss$wtb1to8 <- CW2CB2_regss$wtb
```

```r
CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wtb1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "1-4",]$wtb*(a-
CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wtb1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "5-6",]$wtb*(a-
CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wtb1to8 <- CW2CB2_regss[CW2CB2_regss$Years == "7-8",]$wtb*(a-
mean(CW2CB2_regss$wtb1to8)
```

```
## [1] 0.4817445
```

```r
CW2CB2_regss$wtb1to8 <- CW2CB2_regss$wtb1to8/0.4817444505
summ(CW2CB2_regss$wtb1to8, graph = FALSE)
```

```
##  obs.  mean    median  s.d.   min.   max.
##  6155  1       0       1.618  0      14.577
```

```r
sum(CW2CB2_regss$wtb1to8, graph = FALSE) #Check if the weighting sum up to the sample size we have
```

```
## [1] 6155
```

```r
weightings <- CW2CB2_regss %>% select(ID, wti1to8, wtn1to8, wtb1to8)

# add the weightings to the other datasets
CW3CB2_regss <- CW3CB2_regss %>%
  left_join(weightings, by = "ID")

LCGA_2class <- LCGA_2class %>%
  left_join(weightings, by = "ID")

LCGA_3class <- LCGA_3class %>%
  left_join(weightings, by = "ID")
```

survey designed analysis

2by2 multilevel LCA dataset (BMI)

```r
 # specifying a survey design

CW2CB2_regss$drink <- as.factor(CW2CB2_regss$drink)
CW2CB2_regss$cigsta3 <- as.factor(CW2CB2_regss$cigsta3)

cw2cb2 <- svydesign(id = ~area, strat = ~gor, weights=~wti1to8, data = CW2CB2_regss, nest = TRUE)

summary(svyglm(bmival ~ CB, design = cw2cb2))
```

```
##
## Call:
## svyglm(formula = bmival ~ CB, design = cw2cb2)
##
## Survey design:
## svydesign(id = ~area, strat = ~gor, weights = ~wti1to8, data = CW2CB2_regss,
##     nest = TRUE)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.6815     0.2917  91.468  < 2e-16 ***
```

```
## CB              0.5474      0.2026   2.702  0.00699 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 28.88983)
##
## Number of Fisher Scoring iterations: 2
```

```r
summary(svyglm(bmival ~ CB + AGE + SEX + cigsta3 + drink, design = cw2cb2))
```

```
##
## Call:
## svyglm(formula = bmival ~ CB + AGE + SEX + cigsta3 + drink, design = cw2cb2)
##
## Survey design:
## svydesign(id = ~area, strat = ~gor, weights = ~wti1to8, data = CW2CB2_regss,
##      nest = TRUE)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.286129   0.550766  42.280  < 2e-16 ***
## CB           0.748599   0.199931   3.744 0.000191 ***
## AGE          0.045106   0.005802   7.774 1.78e-14 ***
## SEX         -0.428181   0.185224  -2.312 0.020983 *
## cigsta31    -0.302576   0.280464  -1.079 0.280901
## cigsta32     0.996839   0.300290   3.320 0.000932 ***
## cigsta33     0.185935   0.233103   0.798 0.425249
## drink2       0.372143   0.438018   0.850 0.395732
## drink3       0.733367   0.350124   2.095 0.036441 *
## drink4       1.092460   0.302167   3.615 0.000314 ***
## drink5       1.734902   0.351716   4.933 9.39e-07 ***
## drink6       2.561946   0.467768   5.477 5.39e-08 ***
## drink7       2.768319   0.494494   5.598 2.75e-08 ***
## drink8       1.813137   0.366400   4.949 8.68e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 27.45079)
##
## Number of Fisher Scoring iterations: 2
```

After adjusting for age, sex, smoking, and drinking, subjects in latent class 2 were averagely with 0.748599 kg/m^2 higher BMI compared with subjects in latent class 1.

Compare with results in Stata (they are exactly the same):

```
##
## . use "/home/wangcc-me/Downloads/UKDA-6533-stata11_se/stata11_se/CW2CB2_regss.d
## > ta", clear
##
## .
## . svyset area [pweight = wti1to8], strata(gor)
##
##      pweight: wti1to8
```

```
##              VCE: linearized
##    Single unit: missing
##       Strata 1: gor
##          SU 1: area
##         FPC 1: <zero>
##
## .
## . svy: regress bmival i.CB
## (running regress on estimation sample)
##
## Survey: Linear regression
##
## Number of strata    =          12              Number of obs    =       5,762
## Number of PSUs      =       1,408              Population size   = 5,683.0462
##                                                Design df         =       1,396
##                                                F(   1,   1396)   =        7.30
##                                                Prob > F          =      0.0070
##                                                R-squared         =      0.0023
##
## ------------------------------------------------------------------------------
##              |             Linearized
##       bmival |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
## -------------+----------------------------------------------------------------
##         2.CB |   .5474219   .2025698     2.70   0.007     .1500479    .9447959
##        _cons |   27.22889   .1233456   220.75   0.000     26.98693    27.47085
## ------------------------------------------------------------------------------
##
## . svy: regress bmival i.CB AGE SEX i.cigsta3 i.drink
## (running regress on estimation sample)
##
## Survey: Linear regression
##
## Number of strata    =          12              Number of obs    =       5,759
## Number of PSUs      =       1,408              Population size   = 5,679.1728
##                                                Design df         =       1,396
##                                                F(  13,   1384)   =       39.96
##                                                Prob > F          =      0.0000
##                                                R-squared         =      0.0521
##
## ------------------------------------------------------------------------------
##              |             Linearized
##       bmival |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
## -------------+----------------------------------------------------------------
##         2.CB |   .7485993   .1999234     3.74   0.000     .3564166    1.140782
##          AGE |   .0451059   .0058023     7.77   0.000     .0337237    .0564881
##          SEX |  -.4281809   .1852147    -2.31   0.021    -.7915101   -.0648517
##              |
##      cigsta3 |
##            1 |   -.302576   .2804648    -1.08   0.281    -.8527539    .2476019
##            2 |   .9968394   .3002816     3.32   0.001     .4077876    1.585891
##            3 |   .1859347   .2330859     0.80   0.425    -.2713017    .6431712
##              |
##        drink |
##            2 |   .3721433    .438019     0.85   0.396    -.4871031     1.23139
```

9

```
##             3  |   .7333668    .3501229    2.09   0.036     .0465431      1.42019
##             4  |    1.09246    .3021667    3.62   0.000     .4997101      1.68521
##             5  |   1.734902    .3517108    4.93   0.000     1.044964     2.424841
##             6  |   2.561946    .4677736    5.48   0.000      1.64433     3.479561
##             7  |   2.768319    .4944962    5.60   0.000     1.798283     3.738355
##             8  |   1.813137    .3663827    4.95   0.000     1.094417     2.531857
##               |
##         _cons |   24.03473    .5168893   46.50   0.000     23.02076     25.04869
## ------------------------------------------------------------------------------
```

3by2 multilevel LCA dataset (BMI)

```
# specifying a survey design
CW3CB2_regss$drink <- as.factor(CW3CB2_regss$drink)
CW3CB2_regss$cigsta3 <- as.factor(CW3CB2_regss$cigsta3)


cw3cb2 <- svydesign(id = ~area, strat = ~gor, weights=~wti1to8, data = CW3CB2_regss, nest = TRUE)


summary(svyglm(bmival ~ CB, design = cw3cb2))
```

```
##
## Call:
## svyglm(formula = bmival ~ CB, design = cw3cb2)
##
## Survey design:
## svydesign(id = ~area, strat = ~gor, weights = ~wti1to8, data = CW3CB2_regss,
##     nest = TRUE)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.8067     0.3000  89.344   <2e-16 ***
## CB            0.4309     0.2033   2.119   0.0343 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 28.91208)
##
## Number of Fisher Scoring iterations: 2
```

```
summary(svyglm(bmival ~ CB + AGE + SEX + cigsta3 + drink , design = cw3cb2))
```

```
##
## Call:
## svyglm(formula = bmival ~ CB + AGE + SEX + cigsta3 + drink, design = cw3cb2)
##
## Survey design:
## svydesign(id = ~area, strat = ~gor, weights = ~wti1to8, data = CW3CB2_regss,
##     nest = TRUE)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.489942   0.560335  41.921  < 2e-16 ***
## CB           0.585273   0.204120   2.867 0.004221 **
## AGE          0.045012   0.005796   7.767 1.87e-14 ***
```

```
## SEX          -0.426279    0.184605   -2.309 0.021125 *
## cigsta31      -0.270981    0.284690   -0.952 0.341389
## cigsta32       1.016753    0.298503    3.406 0.000683 ***
## cigsta33       0.184376    0.239308    0.770 0.441199
## drink2         0.329579    0.440043    0.749 0.454040
## drink3         0.727576    0.350977    2.073 0.038410 *
## drink4         1.050976    0.303974    3.457 0.000567 ***
## drink5         1.682150    0.351105    4.791 1.89e-06 ***
## drink6         2.514237    0.473437    5.311 1.33e-07 ***
## drink7         2.707192    0.500577    5.408 7.85e-08 ***
## drink8         1.777640    0.366902    4.845 1.45e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 27.49034)
##
## Number of Fisher Scoring iterations: 2
```

2 classes LCGA dataset (BMI)

```r
# specifying a survey design
LCGA_2class$drink <- as.factor(LCGA_2class$drink)
LCGA_2class$cigsta3 <- as.factor(LCGA_2class$cigsta3)

LCGA_2c <- svydesign(id = ~area, strat = ~gor, weights=~wti1to8, data = LCGA_2class, nest = TRUE)

summary(svyglm(bmival ~ factor(C), design = LCGA_2c))
```

```
##
## Call:
## svyglm(formula = bmival ~ factor(C), design = LCGA_2c)
##
## Survey design:
## svydesign(id = ~area, strat = ~gor, weights = ~wti1to8, data = LCGA_2class,
##     nest = TRUE)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.5888     0.1262 218.666  < 2e-16 ***
## factor(C)2   -0.5381     0.1941  -2.773  0.00565 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 28.89348)
##
## Number of Fisher Scoring iterations: 2
```

```r
summary(svyglm(bmival ~ factor(C) + age + Sex + cigsta3 + drink, design = LCGA_2c))
```

```
##
## Call:
## svyglm(formula = bmival ~ factor(C) + age + Sex + cigsta3 + drink,
##     design = LCGA_2c)
##
```

```
## Survey design:
## svydesign(id = ~area, strat = ~gor, weights = ~wti1to8, data = LCGA_2class,
##     nest = TRUE)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.148633   0.518985  46.531  < 2e-16 ***
## factor(C)2  -0.610432   0.186959  -3.265  0.00113 **
## age          0.047379   0.005803   8.165 8.95e-16 ***
## Sex         -0.430104   0.183833  -2.340  0.01948 *
## cigsta31     0.158571   0.284320   0.558  0.57715
## cigsta32     1.430657   0.300154   4.766 2.13e-06 ***
## cigsta33     0.564804   0.233026   2.424  0.01552 *
## drink2       0.318475   0.446784   0.713  0.47611
## drink3       0.749556   0.352872   2.124  0.03388 *
## drink4       0.914224   0.305128   2.996  0.00280 **
## drink5       1.498078   0.352053   4.255 2.27e-05 ***
## drink6       2.277219   0.463696   4.911 1.05e-06 ***
## drink7       2.462229   0.493897   4.985 7.21e-07 ***
## drink8       1.489875   0.364220   4.091 4.62e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 27.48692)
##
## Number of Fisher Scoring iterations: 2
```

3 classes LCGA dataset (BMI)

```r
# specifying a survey design
LCGA_3class$drink <- as.factor(LCGA_3class$drink)
LCGA_3class$cigsta3 <- as.factor(LCGA_3class$cigsta3)

LCGA_3c <- svydesign(id = ~area, strat = ~gor, weights=~wti1to8, data = LCGA_3class, nest = TRUE)

summary(svyglm(bmival ~ factor(C), design = LCGA_3c))
```

```
##
## Call:
## svyglm(formula = bmival ~ factor(C), design = LCGA_3c)
##
## Survey design:
## svydesign(id = ~area, strat = ~gor, weights = ~wti1to8, data = LCGA_3class,
##     nest = TRUE)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.50674    0.21256 129.407   <2e-16 ***
## factor(C)2  -0.47001    0.37132  -1.266    0.206
## factor(C)3  -0.09078    0.23876  -0.380    0.704
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for gaussian family taken to be 28.94449)
##
## Number of Fisher Scoring iterations: 2
summary(svyglm(bmival ~ factor(C) + age + Sex + cigsta3 + drink, design = LCGA_3c))

##
## Call:
## svyglm(formula = bmival ~ factor(C) + age + Sex + cigsta3 + drink,
##     design = LCGA_3c)
##
## Survey design:
## svydesign(id = ~area, strat = ~gor, weights = ~wti1to8, data = LCGA_3class,
##     nest = TRUE)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.102497   0.517893  46.539  < 2e-16 ***
## factor(C)2  -0.656880   0.364006  -1.805  0.07142 .
## factor(C)3  -0.291928   0.235069  -1.242  0.21455
## age          0.047261   0.005828   8.109 1.38e-15 ***
## Sex         -0.416449   0.183688  -2.267  0.02358 *
## cigsta31     0.202552   0.318284   0.636  0.52466
## cigsta32     1.484134   0.348869   4.254 2.28e-05 ***
## cigsta33     0.620245   0.282956   2.192  0.02859 *
## drink2       0.347662   0.445022   0.781  0.43484
## drink3       0.738233   0.350524   2.106  0.03543 *
## drink4       0.936581   0.303499   3.086  0.00208 **
## drink5       1.522516   0.351407   4.333 1.61e-05 ***
## drink6       2.302737   0.465045   4.952 8.54e-07 ***
## drink7       2.501582   0.490716   5.098 4.06e-07 ***
## drink8       1.503911   0.365952   4.110 4.27e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 27.53784)
##
## Number of Fisher Scoring iterations: 2
```