STAT 332 A1Q4

Yuqi Gu

6/2/2024

```
classurv <- read.csv("~/Desktop/STAT 332/Assignments/A1/classurv.txt", sep="")
library(survey)
```

```
## Warning: package 'survey' was built under R version 4.1.2

## Loading required package: grid

## Loading required package: Matrix

## Loading required package: survival

##
## Attaching package: 'survey'

## The following object is masked from 'package:graphics':
##
##     dotchart
```

(a)

```
classurv$GPA[classurv$GPA == -9] <- NA
N <- 350
svy <- svydesign(ids = ~1, probs = c(nrow(classurv)/N), fpc = rep(N, nrow(classurv)),
                 data=classurv)
summary(svy)
```

```
## Independent Sampling design
## svydesign(ids = ~1, probs = c(nrow(classurv)/N), fpc = rep(N,
##     nrow(classurv)), data = classurv)
## Probabilities:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1629  0.1629  0.1629  0.1629  0.1629  0.1629
## Population size (PSUs): 350
## Data variables:
##  [1] "gender"   "age"      "GPA"      "year"     "height"   "weight"
##  [7] "studyHrs" "sleepHrs" "job"      "textPay"  "reside"
```

```r
GPA_mean <- svymean(~GPA, svy, na.rm = TRUE)
GPA_mean
```

```
##      mean     SE
## GPA 3.006 0.0502
```

```r
dfall <- nrow(subset(classurv, !is.na(GPA)))-1
confint(GPA_mean, "GPA", df=dfall)
```

```
##       2.5 %   97.5 %
## GPA 2.90498 3.107103
```

The point estimate for average GPA of all students is 3.006. The corresponding 95% CI is [2.90498, 3.107103].

(b)

```r
male_students <- classurv[classurv$gender==1,]
n_male <- nrow(male_students)
male_svy <- svydesign(ids = ~1, probs = c(n_male/N), fpc=rep(N,nrow(male_students)),
                     data=male_students)
male_mean <- svymean(~GPA, male_svy, na.rm = TRUE)
male_mean
```

```
##       mean     SE
## GPA 3.0491 0.0676
```

```r
dfmale <- nrow(subset(classurv[classurv$gender==1,], !is.na(GPA)))-1
confint(male_mean, "GPA", df=dfmale)
```

```
##        2.5 %   97.5 %
## GPA 2.908975 3.189286
```

The point estimate for average GPA of all male students is 3.0491. The corresponding 95% CI is [2.908975, 3.189286].

(c)

```r
female_students <- classurv[classurv$gender==2,]
n_female <- nrow(female_students)
female_svy <- svydesign(ids = ~1, probs = c(n_female/N), fpc=rep(N,nrow(female_students)),
                      data=female_students)
female_mean <- svymean(~GPA, female_svy, na.rm = TRUE)
female_mean
```

```
##       mean     SE
## GPA 2.9664 0.0797
```

```
dffemale <- nrow(subset(classurv[classurv$gender==2,], !is.na(GPA)))-1
confint(female_mean, "GPA", df=dffemale)
```

```
##        2.5 %   97.5 %
## GPA 2.801902 3.130898
```

The point estimate for average GPA of all female students is 2.9664. The corresponding 95% CI is [2.801902, 3.130898].

(d)

```
spmale<-svyciprop(~I(gender==1), svy, method="mean", na.rm = TRUE)
spmale
```

```
##                         2.5% 97.5%
## I(gender == 1) 0.404 0.283  0.52
```

```
spfemale<-svyciprop(~I(gender==2), svy, method="mean", na.rm = TRUE)
spfemale
```

```
##                         2.5% 97.5%
## I(gender == 2) 0.596 0.476  0.72
```

The point estimate for the proportion of male students is 0.404. The corresponding 95% CI for the proportion of male students is [0.283, 0.52]. The point estimate for the proportion of female students is 0.596. The corresponding 95% CI for the proportion of female students is [0.476, 0.72].

(e)

```
svyyear1<-svyciprop(~I(year==1), svy, method ="logit", na.rm = TRUE)
svyyear1
```

```
##                         2.5% 97.5%
## I(year == 1) 0.1404 0.0746  0.25
```

```
svyyear2<-svyciprop(~I(year==2), svy, method ="logit", na.rm = TRUE)
svyyear2
```

```
##                         2.5% 97.5%
## I(year == 2) 0.1228 0.0622  0.23
```

```
svyyear3<-svyciprop(~I(year==3), svy, method ="mean", na.rm = TRUE)
svyyear3
```

```
##                       2.5% 97.5%
## I(year == 3) 0.439 0.317  0.56
```

```r
svyyear4<-svyciprop(~I(year==4), svy, method ="mean", na.rm = TRUE)
svyyear4
```

```
##                      2.5% 97.5%
## I(year == 4) 0.263 0.155  0.37
```

```r
svyyear5<-svyciprop(~I(year==5), svy, method ="logit", na.rm = TRUE)
svyyear5
```

```
##                          2.5% 97.5%
## I(year == 5) 0.03509 0.00952  0.12
```

The point estimates for the proportions of students that are freshman/sophomore/junior/senior/other are respectively 0.1404, 0.1228, 0.439, 0.263,0.03509. The corresponding 95% CI for the proportions of students that are freshman/sophomore/junior/senior/other are respectively [0.0746, 0.25], [0.0622, 0.23], [0.317, 0.56], [0.155, 0.37], [0.00952, 0.12]. Note that since the point estimates for the proportions of students that are freshman/sophomore/other are smaller than 0.2, we use logit.

(f)

```r
svyfre <- classurv[classurv$year==1,]
n_svyfreshman <- nrow(svyfre)
freshman_svy <- svydesign(ids = ~1, probs = c(n_svyfreshman/N), fpc=rep(N,nrow(svyfre)),
                     data=svyfre)

svyciprop(~I(gender==1), freshman_svy, method ="logit", na.rm = TRUE)
```

```
##                          2.5% 97.5%
## I(gender == 1) 0.12500 0.00978  0.67
```

```r
svyciprop(~I(gender==2), freshman_svy, method ="logit", na.rm = TRUE)
```

```
##                        2.5% 97.5%
## I(gender == 2) 0.875 0.326  0.99
```

Amongst freshmen, the point estimate for the proportions of male students is 0.125. The point estimate for the proportions of female students is 0.875. Since amongst freshmen, the point estimate for the proportion of male students is smaller than 0.2 and the point estimate for the proportion of female students is larger than 0.8, we use logit. The corresponding 95% CI for the proportion of male students is [0.00978, 0.67] and the corresponding 95% CI for the proportion of female students is [0.326, 0.99].

(g)

```r
svysen <- classurv[classurv$year==4,]
n_svysenior <- nrow(svysen)
senior_svy <- svydesign(ids = ~1, probs = c(n_svysenior/N), fpc=rep(N, nrow(svysen)),
                     data=svysen)

svyciprop(~I(gender==1), senior_svy, method="mean", na.rm = TRUE)
```

```
##                         2.5% 97.5%
## I(gender == 1) 0.600 0.325  0.87
```

```r
svyciprop(~I(gender==2), senior_svy, method="mean", na.rm = TRUE)
```

```
##                         2.5% 97.5%
## I(gender == 2) 0.400 0.125  0.67
```

Amongst seniors, the point estimate for the proportions of male students is 0.6. The corresponding 95% CI for the proportion of male students amongst senior is [0.325, 0.87]. The point estimate for the proportions of female students is 0.4. The corresponding 95% CI for the proportion of female students amongst seniors is [0.125, 0.67].

(h)

```r
topperformers <- classurv[classurv$GPA>=3.2,]
n_top <- nrow(topperformers)
top_svy <- svydesign(ids = ~1, probs = c(n_top/N), fpc=rep(N, nrow(topperformers)),
                     data=topperformers)

svyciprop(~I(gender==1), top_svy, method="mean", na.rm = TRUE)
```

```
##                         2.5% 97.5%
## I(gender == 1) 0.500 0.263  0.74
```

```r
svyciprop(~I(gender==2), top_svy, method="mean", na.rm = TRUE)
```

```
##                         2.5% 97.5%
## I(gender == 2) 0.500 0.263  0.74
```

Amongst the top permormers(i.e., GPA >= 3.2), the point estimate for the proportions of male students is 0.5. The corresponding 95% CI for the proportion of male students amongst senior is [0.263, 0.74]. The point estimate for the proportions of female students is 0.5. The corresponding 95% CI for the proportion of female students amongst seniors is [0.263, 0.74].