A2Q4

Yuqi Gu

6/14/2024

```
syc <- read.csv("~/Desktop/STAT 332/Assignments/A2/syc.txt")
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
```

```
library(car)
```

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

```
## Loading required package: carData
```

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

```
syc$age[syc$age == 99] <- NA
syc$numarr[syc$numarr == 99] <- NA
syc$race <- recode(syc$race, 'c(1)="white"; c(2)="black";
                   c(3)="Asian/Pacific Islander"; c(4)="Native"; c(5)="other";
                   else=NA')
syc$sex <- recode(syc$sex, 'c(1)="male"; c(2)="female"; else=NA')
syc$prviol <- recode(syc$prviol, 'c(1)="yes"; c(0)="no"; else=NA')
syc$everdrug <- recode(syc$everdrug, 'c(1)="yes"; c(0)="no"; else=NA')
```

```
Nh <- c(2724, 3192, 4107, 2705, 3504, 376, 56, 528, 624, 520, 672, 384, 744, 847, 824, 1848)
Nh <- data.frame(stratum=seq(1,16), Nh=as.vector(Nh))

nh <- table(syc$stratum) # would not be used
nh <- data.frame(stratum=names(nh), nh=as.vector(nh)) # would not be used

syc.stsrs <- merge(syc, Nh, by = "stratum")
syc.stsrs <- merge(syc.stsrs, nh, by = "stratum") # would not be used
syc.stsrs$pik <- syc.stsrs$nh/syc.stsrs$Nh # would not be used

syc.svy <- svydesign(ids=~1, strata = ~stratum,  weights=~finalwt, fpc=~Nh, data = syc.stsrs)
```

(a)

```
tmp1 <- svymean(~age, syc.svy, na.rm=T)
tmp1
```

```
##        mean      SE
## age 16.639 0.0312
```

```
confint(tmp1, names(tmp1), df=degf(syc.svy))
```

```
##        2.5 %    97.5 %
## age 16.57815 16.70043
```

The point estimate of the average age of juveniles/young adults in custody is 16.639. The corresponding 95% CI is [16.57815, 16.70043].

(b)

```
tmp2 <- svymean(~numarr, syc.svy, na.rm=T)
tmp2
```

```
##           mean      SE
## numarr 8.9297 0.2481
```

```
numarrNA <- sum(is.na (syc.svy$variable[, "numarr"]))
confint(tmp2, names(tmp2), df=nrow(syc)-numarrNA-16)
```

```
##          2.5 % 97.5 %
## numarr 8.443168 9.41632
```

The point estimate of the mean number of prior arrests is 8.9297. The corresponding 95% CI is [8.443168, 9.41632].

(c)

```
svyciprop(~I(everdrug=="yes"), syc.svy, method="logit", na.rm = TRUE)
```

```
##                           2.5% 97.5%
## I(everdrug == "yes") 0.828 0.811   0.84
```

The point estimate of the proportion of juveniles/young adults in custody that have used illegal drugs is 0.828. The corresponding 95% CI is [0.811, 0.84].

(d)

```
svyciprop(~I(prviol=="yes"), syc.svy, method="mean", na.rm = TRUE)
```

```
##                          2.5% 97.5%
## I(prviol == "yes") 0.355 0.335   0.37
```

The point estimate of the proportion of juveniles/young adults in custody that were previously arrested for a violent crime is 0.355. The corresponding 95% CI is [0.335, 0.37].

(e)

```
svyciprop(~I(interaction(everdrug,prviol) == "yes.yes"), syc.svy, method="mean", na.rm = TRUE)
```

```
##                                             2.5% 97.5%
## I(interaction(everdrug, prviol) == "yes.yes") 0.315 0.296   0.33
```

The point estimate of the proportion of juveniles/young adults in custody that have used illegal drugs and were previously arrested for a violent crime is 0.315. The corresponding 95% CI is [0.296, 0.33].

(f)

```
svyciprop(~I(sex == "male"), syc.svy, method="logit", na.rm = TRUE)
```

```
##                      2.5% 97.5%
## I(sex == "male") 0.931 0.920   0.94
```

The point estimate of the proportion of males amongst juveniles/young adults in custody is 0.931. The corresponding 95% CI is [0.920, 0.94].

(g)

```
svyciprop(~I(race == "black"), syc.svy, method="mean", na.rm = TRUE)
```

```
##                        2.5% 97.5%
## I(race == "black") 0.415 0.395   0.44
```

The point estimate of the proportion of African Americans amongst juveniles/young adults in custody is 0.415. The corresponding 95% CI is [0.395, 0.44].

(h)

```
svyciprop(~I(interaction(race, sex) == "black.male"), syc.svy, method="mean", na.rm = TRUE)
```

```
##                                        2.5% 97.5%
## I(interaction(race, sex) == "black.male") 0.398 0.377  0.42
```

The point estimate of the proportion of African American males amongst juveniles/young adults in custody is 0.398. The corresponding 95% CI is [0.377, 0.42].

(i)

From census data, we know that African-American males are about 7.6% of the youth population in the U.S. On the contrary, in part(h), the point estimate of the proportion of African American males amongst juveniles/young adults in custody is 0.398, which is much higher than 7.6%. The bias is 7.6% - 39.8% = -0.322%. When it comes to youth in custody, there is racial bias that there are many African-American males in custody.