SURVMETH 745 HW8

Stacey Frank & Chendi Zhao

4/4/2022

# Q1.

## OUTCOME Ineligible Nonresponse Response Unknown Eligible Sum  
## OUTCOME1   
## 201 0 0 35823 0 35823  
## 203 0 0 9998 0 9998  
## 213 0 107 0 0 107  
## 215 0 2062 0 0 2062  
## 216 0 0 0 3001 3001  
## 217 0 396 0 0 396  
## 218 0 8823 0 0 8823  
## 219 0 1727 0 0 1727  
## 299 8853 0 0 0 8853  
## Sum 8853 13115 45821 3001 70790

# Q2.

e=(13115+45821)/(70790-3001)  
RR3<-35823/(45821+13115+e\*3001)  
RR3

The response rate for these data, as calculated by the AAPOR Response Rate 3 formula, is 58.2%.

# Q3.

In each adjustment class b: The unknown eligibility adjustment for sample units is , where is the base weight The final adjusted weight is

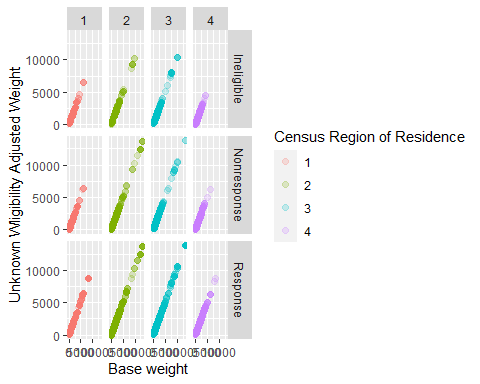
table(dat$CENREG,dat$OUTCOME,exclude = NULL)

##   
## Ineligible Nonresponse Response Unknown Eligible  
## 1 1723 2888 7562 922  
## 2 1730 2558 9639 493  
## 3 3002 4351 16044 886  
## 4 2398 3318 12576 700

# Group data by CENREG and OUTCOME; sum weights within those groupings  
dat2<-dat %>%  
 group\_by(CENREG, OUTCOME) %>%  
 summarise(base=sum(WTIA\_PD))  
  
# Calculate a1b for each of the 4 classes  
w1=sum(dat2$base[1:4])/sum(dat2$base[1:3])  
w2=sum(dat2$base[5:8])/sum(dat2$base[5:7])  
w3=sum(dat2$base[9:12])/sum(dat2$base[9:11])  
w4=sum(dat2$base[13:16])/sum(dat2$base[13:15])  
rbind(w1,w2,w3,w4)

## [,1]  
## w1 1.066633  
## w2 1.028876  
## w3 1.033133  
## w4 1.032186

dat3<-dat %>%  
 group\_by(CENREG, OUTCOME) %>%  
 summarise(base=WTIA\_PD,  
 CTSTAT3)  
  
# Calculate unknown eligibility adjusted weights  
dat3$WT\_Eligible<-case\_when(  
 dat3$CENREG==1&!dat3$OUTCOME=="Unknown Eligible"~dat3$base\*w1,  
 dat3$CENREG==2&!dat3$OUTCOME=="Unknown Eligible"~dat3$base\*w2,  
 dat3$CENREG==3&!dat3$OUTCOME=="Unknown Eligible"~dat3$base\*w3,  
 dat3$CENREG==4&!dat3$OUTCOME=="Unknown Eligible"~dat3$base\*w4,  
 dat3$OUTCOME=="Unknown Eligible"~dat3$base\*0,  
 TRUE~as.numeric(NA))  
  
#Drop unknown eligibility cases from data frame  
dat4<-dat3%>%filter(!OUTCOME=="Unknown Eligible")  
  
dat4%>%ggplot(aes(base,WT\_Eligible,color=factor(CENREG)))+  
 geom\_point(alpha=0.2, size=2) +  
 facet\_grid(vars(OUTCOME),vars(CENREG))+  
 labs(x="Base weight",  
 y="Unknown Wligibility Adjusted Weight",  
 color="Census Region of Residence")



# Q4.

dat5<-dat%>%filter(OUTCOME=="Nonresponse")  
  
unwt.mean=mean(dat5$CTSTAT3,na.rm=T)  
  
wt.dsn <- svydesign(~1, strata = NULL, weights = ~WTIA\_PD, data = dat5)  
wt.mean <- svymean(~CTSTAT3, wt.dsn, na.rm=TRUE)  
  
unwt.mean;wt.mean

## [1] 4.858154

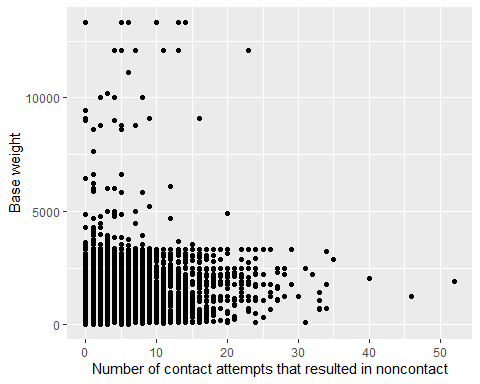
## mean SE  
## CTSTAT3 4.9125 0.0472

dat5%>%ggplot(aes(dat5$CTSTAT3,dat5$WTIA\_PD))+  
 geom\_point()+  
 labs(x="Number of contact attempts that resulted in noncontact",  
 y="Base weight")

## Warning: Use of `dat5$CTSTAT3` is discouraged. Use `CTSTAT3` instead.

## Warning: Use of `dat5$WTIA\_PD` is discouraged. Use `WTIA\_PD` instead.

## Warning: Removed 122 rows containing missing values (geom\_point).



The mean of CTSTAT3 among nonrespondents without base weights is 4.86, while after using base weight it is 4.91. The mean value of CTSTAT3 is inflated slightly when using the base weights because, as can be seen from the scatterplot of baseweight and CTSTAT3 above, the cases with base weights above 10,000 mostly required more than 5 contact attempts before they were classified as noncontacts. The large base weights for those cases are driving the weighted mean value to be highter.

However, it should be noted that the NHIS description file says “not every analysis involving the paradata File will entail making population inferences. For example, a data user may want to describe the level of effort (number of contact attempts) expended by NHIS interviewers across a wide range of final case dispositions (complete interview, partial interview, noncontact, refusal, etc.). Use of WTIA\_PD would not be necessary in this situation.” The number of contact attempts it took to reach a house is not intrinsically related to that house’s selection probability, so it doesn’t really make sense to use the weights when analyzing this variable.

# Q5a.

# Q5b.

# Q5c.

# Q6.