HW 3

Caitlin Bolz

3/10/2021

# Question 1

Fitted a guide classification forest to the CE data. Used ceclass.dsc as the data description file and cedata.txt as the data file. Input file is cein.txt, output file is ceout.txt, and predicted class and probability file is cepred.txt

# Question 2

z = read.table("cedata.txt", header = T)  
w <- z$FINLWT21 ### sampling weights  
zclass <- read.table("cepred.txt",header=TRUE)  
probmissing <- zclass[,2] ### estimated P(INTRDVX\_ = C)  
  
p <- 1-probmissing ### estimated P(INTRDVX is nonmissing)  
  
group <- !is.na(z$INTRDVX) ### group of nonmissing INTRDVXobs  
  
ipw <- sum(w[group]\*z$INTRDVX[group]/p[group])/sum(w[group]/p[group])

ipw

## [1] 5010.44

# Question 3

Fitted a guide regression forest to the CE data. Used cereg.dsc as the data description file and cedata.txt as the data file. Input file is regin.txt, output file is regout.txt, and predicted class and probability file is regpred.txt

# Question 4

zreg <- read.table("regpred",header=TRUE)  
  
yhat <- zreg$predicted  
  
imputed <- (sum(w[group]\*z$INTRDVX[group])+sum(w[!group]\*yhat[!group]))/sum(w)  
  
simple <- sum(w[group]\*z$INTRDVX[group])/sum(w[group])

imputed

## [1] 4704.046

# Question 5

## Question 1 Input File: cein.txt

## Question 1 Output File: ceout.txt

## Question 3 Input File: regin.txt

## Question 3 Output File: regout.txt