Homework 3

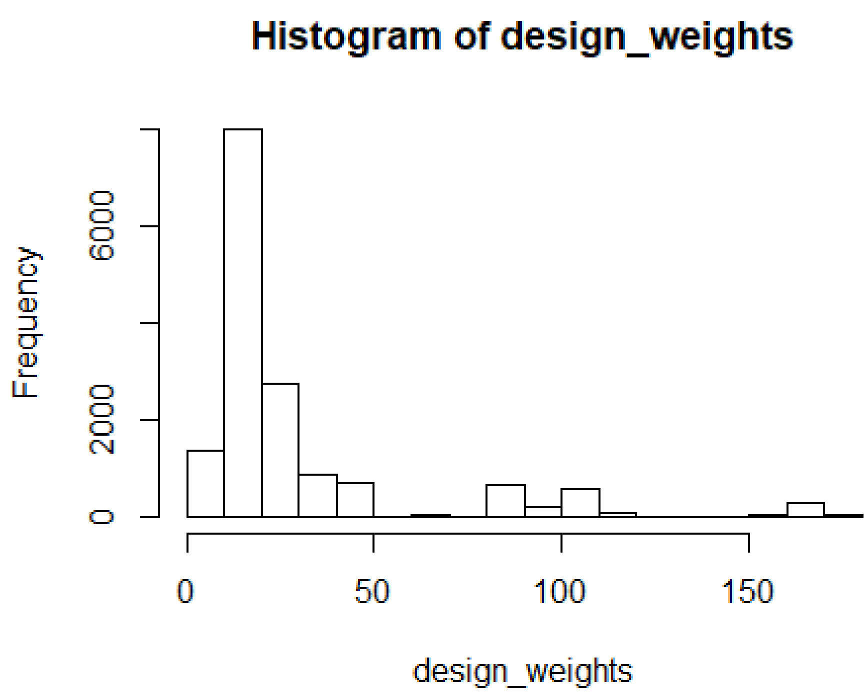
Xinyi Lin

9/26/2019

Question 1

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.083  14.269  17.445  30.205  29.664 178.296
```

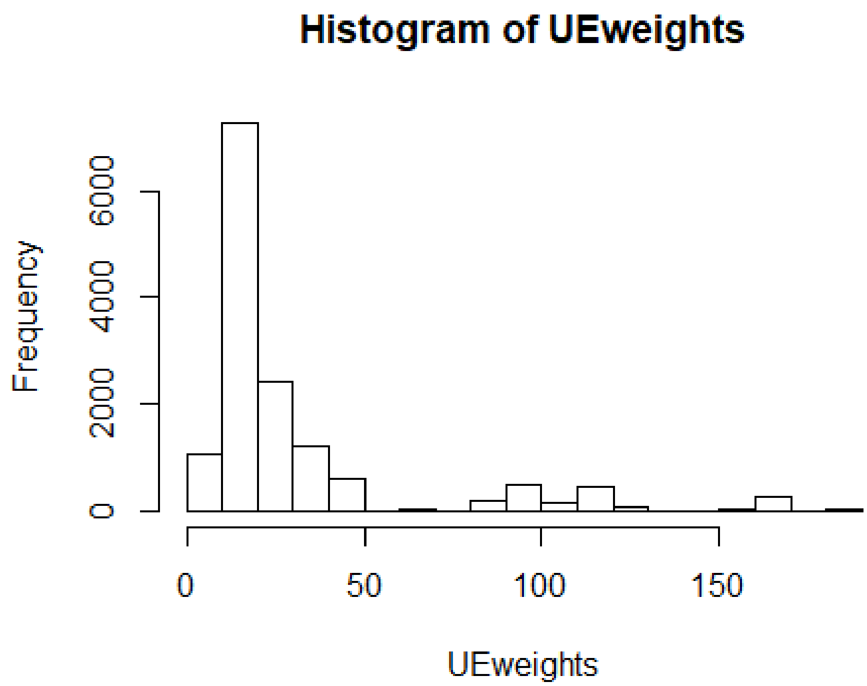
The minimum, quartiles and maximum of design weights are shown above, the histogram is shown below.



Question 2

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.169  14.428  17.657  30.587  29.800 181.725
```

The minimum, quartiles and maximum of adjusted design weights by unknown eligibility are shown above, the histogram is shown below.

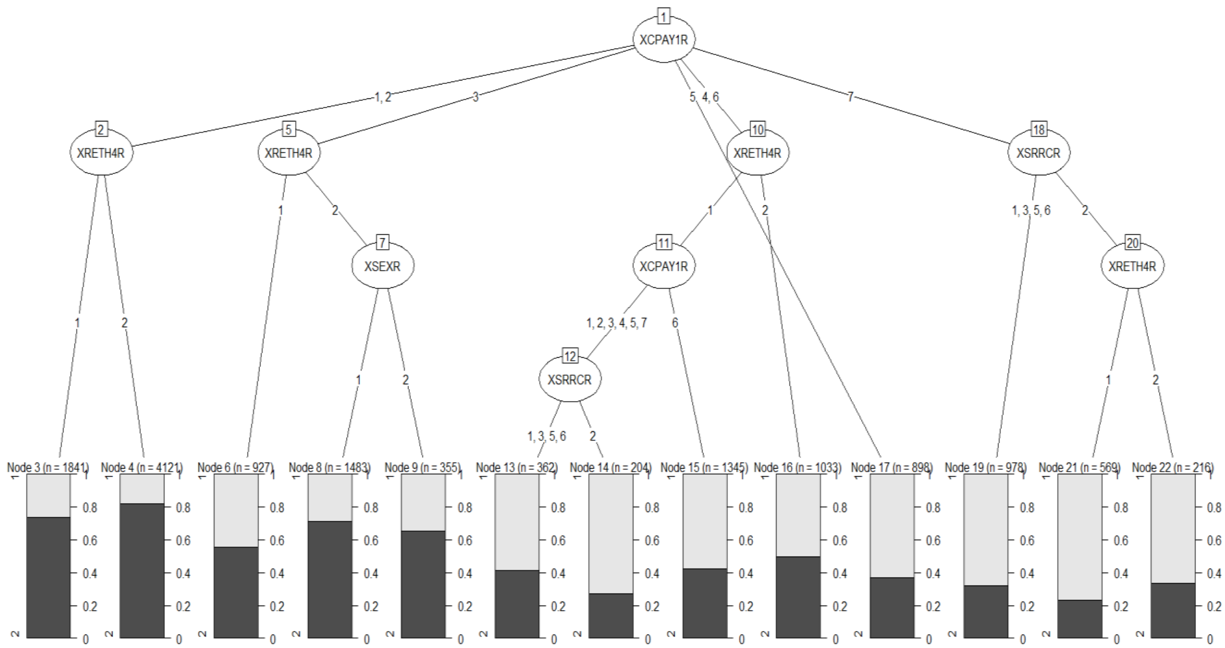


Question 3

In order to apply adjustments for nonresponse to the weights using the CHAID method, I first use the CHAID package to seperate data into different cells, then calculate $w_{nr,c,nw} = \frac{1}{rrate_{c,nw}}, rrate_{c,nw} = \frac{n_{r,c}}{n_c}$ for each cell, final weights is $w_{des}w_{eli}w_{nr,c,nw}$.

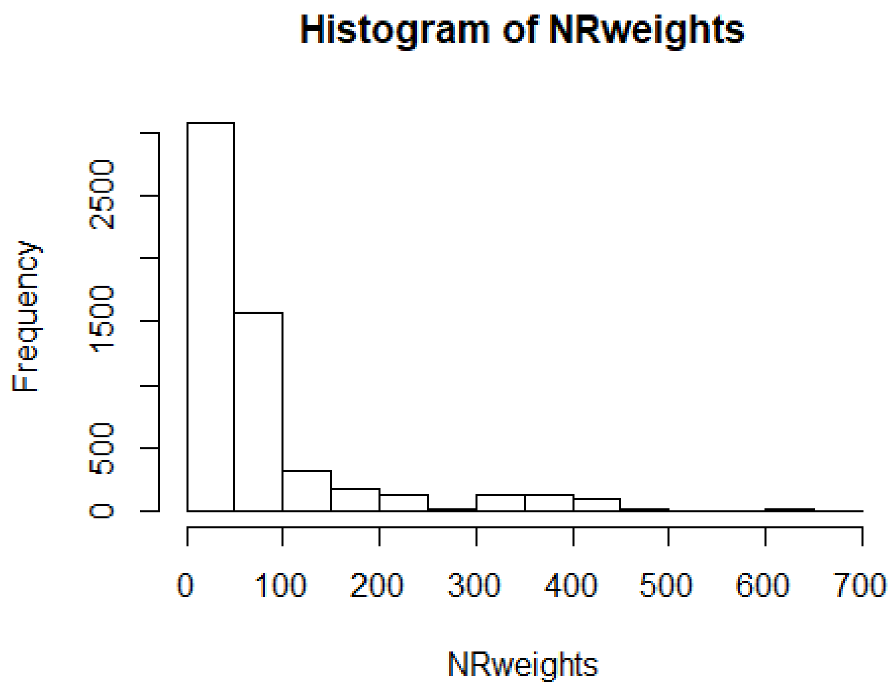
```
##
## Model formula:
## RESPSTAT2 ~ XSEXR + XSRRCR + XCPAY1R + XRETH4R
##
## Fitted party:
## [1] root
## |   [2] XCPAY1R in 1, 2
## |   |   [3] XRETH4R in 1: 2 (n = 1841, err = 26.6%)
## |   |   [4] XRETH4R in 2: 2 (n = 4121, err = 18.4%)
## |   [5] XCPAY1R in 3
## |   |   [6] XRETH4R in 1: 2 (n = 927, err = 44.7%)
## |   |   [7] XRETH4R in 2
## |   |   |   [8] XSEXR in 1: 2 (n = 1483, err = 29.1%)
## |   |   |   [9] XSEXR in 2: 2 (n = 355, err = 34.6%)
## |   [10] XCPAY1R in 4, 6
## |   |   [11] XRETH4R in 1
## |   |   |   [12] XCPAY1R in 1, 2, 3, 4, 5, 7
## |   |   |   |   [13] XSRRCR in 1, 3, 5, 6: 1 (n = 362, err = 41.4%)
## |   |   |   |   [14] XSRRCR in 2: 1 (n = 204, err = 27.0%)
```

```
## |   |   |   [15] XCPAY1R in 6: 1 (n = 1345, err = 42.2%)
## |   |   [16] XRETH4R in 2: 1 (n = 1033, err = 49.8%)
## |   [17] XCPAY1R in 5: 1 (n = 898, err = 36.9%)
## |   [18] XCPAY1R in 7
## |   |   [19] XSRRCR in 1, 3, 5, 6: 1 (n = 978, err = 32.1%)
## |   |   [20] XSRRCR in 2
## |   |   |   [21] XRETH4R in 1: 1 (n = 569, err = 23.4%)
## |   |   |   [22] XRETH4R in 2: 1 (n = 216, err = 33.3%)
##
## Number of inner nodes:     9
## Number of terminal nodes: 13
```




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.56   18.87   33.34   77.70   87.74  663.24
```

The minimum, quartiles and maximum of adjusted design weights by unknown eligibility and nonresponse are shown above, the histogram is shown below.



Appendix

```
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(CHAID)
## Question 1
data = read.csv("./HW3.csv")
p1_data = data %>%
  mutate(design_weights = NSTRAT/NSAMP)
summary(p1_data$design_weights)
design_weights = p1_data$design_weights
hist(design_weights)
## Question 2
UEwei_data = data %>%
  group_by(STRATUM, RESPSTAT2) %>%
  summarize(total = sum(NSAMP)) %>%
  mutate(RESPSTAT2 = str_c("respSTAT_", RESPSTAT2)) %>%
  spread(key = RESPSTAT2, value = total) %>%
  mutate(respSTAT_3 = ifelse(is.na(respSTAT_3), 0, respSTAT_3),
         respSTAT_4 = ifelse(is.na(respSTAT_4), 0, respSTAT_4)) %>%
  mutate(total_num = respSTAT_1 + respSTAT_2 + respSTAT_3 + respSTAT_4,
         UEweights = total_num/(respSTAT_1 + respSTAT_2 + respSTAT_3)) %>%
  select(STRATUM, UEweights)

p2_data = merge(p1_data, UEwei_data) %>%
  mutate(UEweights = UEweights*design_weights) %>%
  filter(RESPSTAT2 %in% c(1,2))
```

```r
summary(p2_data$UEweights)
UEweights = p2_data$UEweights
hist(UEweights)
## Question 3
set.seed(123)
p3_data = data %>%
  filter(RESPSTAT2 %in% c(1,2)) %>%
  mutate(RESPSTAT2 = as.factor(RESPSTAT2),
         XSEXR = as.factor(XSEXR),
         XSRRCR = as.factor(XSRRCR),
         XCPAY1R = as.factor(XCPAY1R),
         XRETH4R = as.factor(XRETH4R))

chaid_data = chaid(RESPSTAT2 ~ XSEXR + XSRRCR + XCPAY1R + XRETH4R, data =
p3_data)
print(chaid_data)
plot(chaid_data)
NR_data = p2_data %>%
  mutate(id = 1:nrow(p2_data))
# cell1
cell1 = NR_data %>%
  filter(XCPAY1R %in% c(1,2) & XRETH4R==1) %>%
  select(id, RESPSTAT2)
num2 = sum(cell1$RESPSTAT2)-nrow(cell1)
cell1_w = nrow(cell1)/(nrow(cell1)-num2)
cell1 = cell1 %>%
  mutate(NRweights = cell1_w)
# cell2
cell2 = NR_data %>%
  filter(XCPAY1R %in% c(1,2) & XRETH4R==2) %>%
  select(id, RESPSTAT2)
num2 = sum(cell2$RESPSTAT2)-nrow(cell2)
cell2_w = nrow(cell2)/(nrow(cell2)-num2)
cell2 = cell2 %>%
  mutate(NRweights = cell2_w)
# cell3
cell3 = NR_data %>%
  filter(XCPAY1R==3 & XRETH4R==1) %>%
  select(id, RESPSTAT2)
num2 = sum(cell3$RESPSTAT2)-nrow(cell3)
cell3_w = nrow(cell3)/(nrow(cell3)-num2)
cell3 = cell3 %>%
  mutate(NRweights = cell3_w)
# cell4
cell4 = NR_data %>%
  filter(XCPAY1R==3 & XRETH4R==2 & XSEXR==1) %>%
  select(id, RESPSTAT2)
num2 = sum(cell4$RESPSTAT2)-nrow(cell4)
cell4_w = nrow(cell4)/(nrow(cell4)-num2)
```

```
cell4 = cell4 %>%
  mutate(NRweights = cell4_w)
# cell5
cell5 = NR_data %>%
  filter(XCPAY1R==3 & XRETH4R==2 & XSEXR==2) %>%
  select(id, RESPSTAT2)
num2 = sum(cell5$RESPSTAT2)-nrow(cell5)
cell5_w = nrow(cell5)/(nrow(cell5)-num2)
cell5 = cell5 %>%
  mutate(NRweights = cell5_w)
# cell6
cell6 = NR_data %>%
  filter(XCPAY1R %in% c(4,6) & XRETH4R==1 & XCPAY1R %in% c(1,2,3,4,5,7) &
XSRRCR %in% c(1,3,5,6)) %>%
  select(id, RESPSTAT2)
num2 = sum(cell6$RESPSTAT2)-nrow(cell6)
cell6_w = nrow(cell6)/(nrow(cell6)-num2)
cell6 = cell6 %>%
  mutate(NRweights = cell6_w)
# cell7
cell7 = NR_data %>%
  filter(XCPAY1R %in% c(4,6) & XRETH4R==1 & XCPAY1R %in% c(1,2,3,4,5,7) &
XSRRCR==2) %>%
  select(id, RESPSTAT2)
num2 = sum(cell7$RESPSTAT2)-nrow(cell7)
cell7_w = nrow(cell7)/(nrow(cell7)-num2)
cell7 = cell7 %>%
  mutate(NRweights = cell7_w)
# cell8
cell8 = NR_data %>%
  filter(XCPAY1R %in% c(4,6) & XRETH4R==1 & XCPAY1R==6) %>%
  select(id, RESPSTAT2)
num2 = sum(cell8$RESPSTAT2)-nrow(cell8)
cell8_w = nrow(cell8)/(nrow(cell8)-num2)
cell8 = cell8 %>%
  mutate(NRweights = cell8_w)
# cell9
cell9 = NR_data %>%
  filter(XCPAY1R %in% c(4,6) & XRETH4R==2) %>%
  select(id, RESPSTAT2)
num2 = sum(cell9$RESPSTAT2)-nrow(cell9)
cell9_w = nrow(cell9)/(nrow(cell9)-num2)
cell9 = cell9 %>%
  mutate(NRweights = cell9_w)
# cell10
cell10 = NR_data %>%
  filter(XCPAY1R==5) %>%
  select(id, RESPSTAT2)
num2 = sum(cell10$RESPSTAT2)-nrow(cell10)
cell10_w = nrow(cell10)/(nrow(cell10)-num2)
```

```r
cell10 = cell10 %>%
  mutate(NRweights = cell10_w)
# cell11
cell11 = NR_data %>%
  filter(XCPAY1R==7 & XSRRCR %in% c(1,3,5,6)) %>%
  select(id, RESPSTAT2)
num2 = sum(cell11$RESPSTAT2)-nrow(cell11)
cell11_w = nrow(cell11)/(nrow(cell11)-num2)
cell11 = cell11 %>%
  mutate(NRweights = cell11_w)
# cell12
cell12 = NR_data %>%
  filter(XCPAY1R==7 & XSRRCR==2 & XRETH4R==1) %>%
  select(id, RESPSTAT2)
num2 = sum(cell12$RESPSTAT2)-nrow(cell12)
cell12_w = nrow(cell12)/(nrow(cell12)-num2)
cell12 = cell12 %>%
  mutate(NRweights = cell12_w)
# cell13
cell13 = NR_data %>%
  filter(XCPAY1R==7 & XSRRCR==2 & XRETH4R==2) %>%
  select(id, RESPSTAT2)
num2 = sum(cell13$RESPSTAT2)-nrow(cell13)
cell13_w = nrow(cell13)/(nrow(cell13)-num2)
cell13 = cell13 %>%
  mutate(NRweights = cell13_w)
cell_data = rbind(cell1, cell2, cell3, cell4, cell5, cell6, cell7, cell8,
cell9, cell10, cell11, cell12, cell13)
p3_data = merge(NR_data, cell_data) %>%
  mutate(NRweights = NRweights*UEweights) %>%
  filter(RESPSTAT2==1)
summary(p3_data$NRweights)
NRweights = p3_data$NRweights
hist(NRweights)
```