DRP Presentation

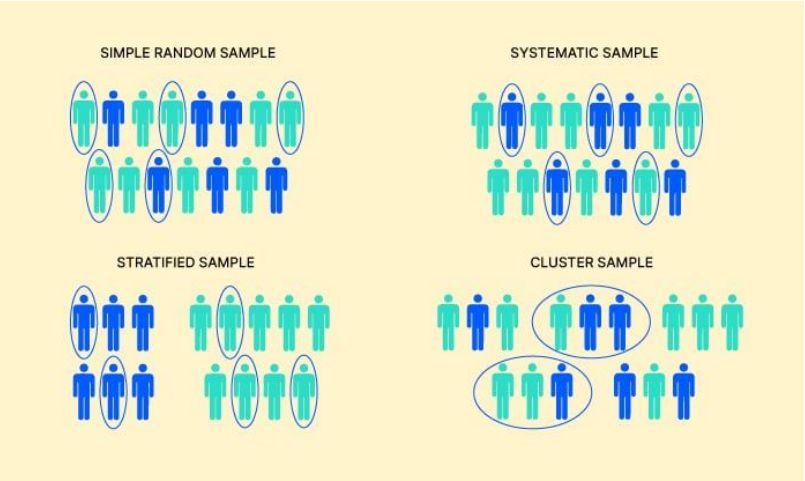
Mentee: Mekias Kebede
Mentor: Jess Kunke

Overview of what I've done

- Survey statistics
  - Sampling Methods
  - Estimators
- R programming
  - Functions/commands
  - Conditionals
  - Data visualization
- Communicating concepts
  - Making concepts understandable
  - Not assuming any prior knowledge

Sampling Methods

- Simple random sampling (WOR)
  - Equal Probability Selection Method
- Auxiliary Information
  - Uses additionally known information to adjust sampling technique (probability proportional to size)
- Cluster sampling
  - Minimizes costs to sample
- Many other methods of sampling
  - Stratified sampling, Systematic sampling, etc



Horvitz-Thompson Estimators

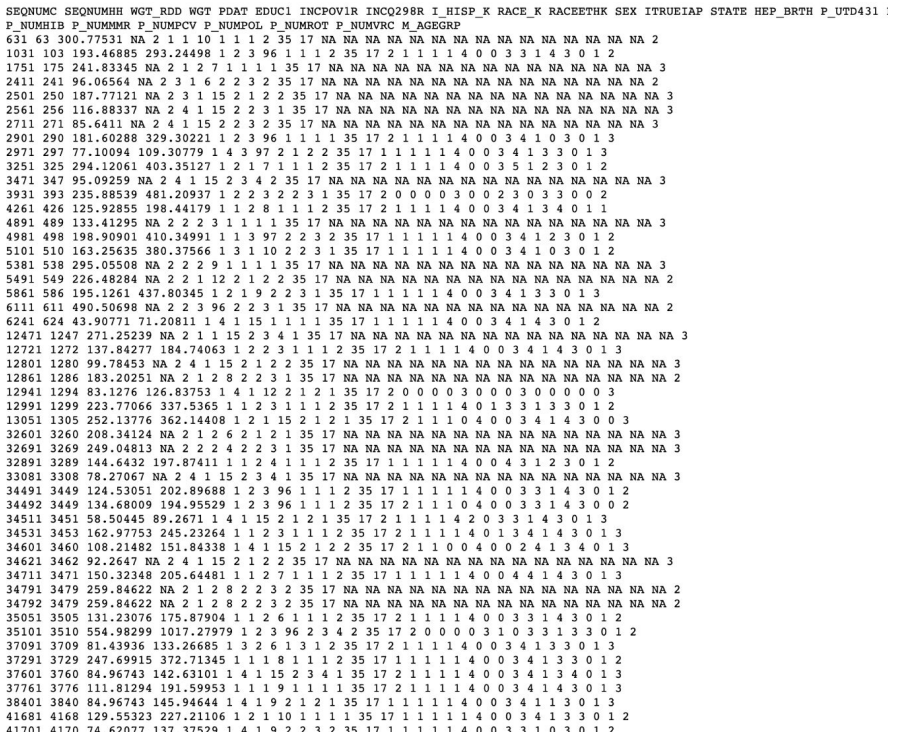
$$\hat{Y} = \sum_{i \in s} \frac{y_i}{\pi_i}$$

Another way to represent this for simple random sampling is to factor out our inclusion probability from the sum and denote π_i as $\frac{n}{N}$ where n is our sample size and N is our total population size.

$$\hat{Y} = \frac{N}{n} \sum_{i \in s} y_i$$

Chicago Immunization Data

- Part of National Immunization Survey
  - CDC conducted survey
  - Immunization status of Children
- For my project:
  - “Pretended” survey was sampling my frame
  - Explored sampling methods.
- Advantages
  - Logistical reasons prevent me from conducting surveys
  - We have the true “answers” to compare estimates with



```
SEQNUMC SEQNUMHH WGT_RDD WGT PDAT EDUC1 INCPOV1R INCQ298R I_HISP_K RACE_K RACEETHK SEX ITRUEIAP STATE HEP_BRTH P_UTD431
P_NUMHIB P_NUMMMR P_NUMPCV P_NUMPOL P_NUMROT P_NUMVRC M_AGEGRP
631 63 300.77531 NA 2 1 1 10 1 1 1 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA 2
1031 103 193.46885 293.24498 1 2 3 96 1 1 1 2 35 17 2 1 1 1 1 4 0 0 3 3 1 4 3 0 1 2
1751 175 241.83345 NA 2 1 2 7 1 1 1 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
2411 241 96.06564 NA 2 3 1 6 2 2 3 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 2
2501 250 187.77121 NA 2 3 1 15 2 1 2 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
2561 256 116.88337 NA 2 4 1 15 2 2 3 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
2711 271 85.6411 NA 2 4 1 15 2 2 3 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
2901 290 181.60288 329.30221 1 2 3 96 1 1 1 1 35 17 2 1 1 1 1 4 0 0 3 4 1 0 3 0 1 3
2971 297 77.10094 109.30779 1 4 3 97 2 1 2 2 35 17 1 1 1 1 1 4 0 0 3 4 1 3 3 0 1 3
3251 325 294.12061 403.35127 1 2 1 7 1 1 1 2 35 17 2 1 1 1 1 4 0 0 3 5 1 2 3 0 1 2
3471 347 95.09259 NA 2 4 1 15 2 3 4 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
3931 393 235.88539 481.20937 1 2 2 3 2 2 3 1 35 17 2 0 0 0 0 3 0 0 2 3 0 3 3 0 0 2
4261 426 125.92855 198.44179 1 1 2 8 1 1 1 2 35 17 2 1 1 1 1 4 0 0 3 4 1 3 4 0 1 1
4891 489 133.41295 NA 2 2 2 3 1 1 1 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
4981 498 198.90901 410.34991 1 1 3 97 2 2 3 2 35 17 1 1 1 1 1 4 0 0 3 4 1 2 3 0 1 2
5101 510 163.25635 380.37566 1 3 1 10 2 2 3 1 35 17 1 1 1 1 1 4 0 0 3 4 1 0 3 0 1 2
5381 538 295.05508 NA 2 2 2 9 1 1 1 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
5491 549 226.48284 NA 2 2 1 12 2 1 2 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 2
5861 586 195.1261 437.80345 1 2 1 9 2 2 3 1 35 17 1 1 1 1 1 4 0 0 3 4 1 3 3 0 1 3
6111 611 490.50698 NA 2 2 3 96 2 2 3 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 2
6241 624 43.90771 71.20811 1 4 1 15 1 1 1 1 35 17 1 1 1 1 1 4 0 0 3 4 1 4 3 0 1 2
12471 1247 271.25239 NA 2 1 1 15 2 3 4 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
12721 1272 137.84277 184.74063 1 2 2 3 1 1 1 2 35 17 2 1 1 1 1 4 0 0 3 4 1 4 3 0 1 3
12801 1280 99.78453 NA 2 4 1 15 2 1 2 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
12861 1286 183.20251 NA 2 1 2 8 2 2 3 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 2
12941 1294 83.1276 126.83753 1 4 1 12 2 1 2 1 35 17 2 0 0 0 0 3 0 0 0 3 0 0 0 0 0 3
12991 1299 223.77066 337.5365 1 1 2 3 1 1 1 2 35 17 2 1 1 1 1 4 0 1 3 3 1 3 3 0 1 2
13051 1305 252.13776 362.14408 1 2 1 15 2 1 2 1 35 17 2 1 1 1 0 4 0 0 3 4 1 4 3 0 0 3
32601 3260 208.34124 NA 2 1 2 6 2 1 2 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
32691 3269 249.04813 NA 2 2 2 4 2 2 3 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
32891 3289 144.6432 197.87411 1 1 2 4 1 1 1 2 35 17 1 1 1 1 1 4 0 0 4 3 1 2 3 0 1 2
33081 3308 78.27067 NA 2 4 1 15 2 3 4 1 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
34491 3449 124.53051 202.89688 1 2 3 96 1 1 1 2 35 17 1 1 1 1 1 4 0 0 3 3 1 4 3 0 1 2
34492 3449 134.68009 194.95529 1 2 3 96 1 1 1 2 35 17 2 1 1 1 0 4 0 0 3 3 1 4 3 0 0 2
34511 3451 58.50445 89.2671 1 4 1 15 2 1 2 1 35 17 2 1 1 1 1 4 2 0 3 3 1 4 3 0 1 3
34531 3453 162.97753 245.23264 1 1 2 3 1 1 1 2 35 17 2 1 1 1 1 4 0 1 3 4 1 4 3 0 1 3
34601 3460 108.21482 151.84338 1 4 1 15 2 1 2 2 35 17 2 1 1 0 0 4 0 0 2 4 1 3 4 0 1 3
34621 3462 92.2647 NA 2 4 1 15 2 1 2 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
34711 3471 150.32348 205.64481 1 1 2 7 1 1 1 2 35 17 1 1 1 1 1 4 0 0 4 4 1 4 3 0 1 3
34791 3479 259.84622 NA 2 1 2 8 2 2 3 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 2
34792 3479 259.84622 NA 2 1 2 8 2 2 3 2 35 17 NA NA NA NA NA NA NA NA NA NA NA NA NA NA 2
35051 3505 131.23076 175.87904 1 1 2 6 1 1 1 2 35 17 2 1 1 1 1 4 0 0 3 3 1 4 3 0 1 2
35101 3510 554.98299 1017.27979 1 2 3 96 2 3 4 2 35 17 2 0 0 0 0 3 1 0 3 3 1 3 3 0 1 2
37091 3709 81.43936 133.26685 1 3 2 6 1 3 1 2 35 17 2 1 1 1 1 4 0 0 3 4 1 3 3 0 1 3
37291 3729 247.69915 372.71345 1 1 1 8 1 1 1 2 35 17 1 1 1 1 1 4 0 0 3 4 1 3 3 0 1 2
37601 3760 84.96743 142.63101 1 4 1 15 2 3 4 1 35 17 2 1 1 1 1 4 0 0 3 4 1 3 4 0 1 3
37761 3776 111.81294 191.59953 1 1 1 9 1 1 1 1 35 17 2 1 1 1 1 4 0 0 3 4 1 4 3 0 1 3
38401 3840 84.96743 145.94644 1 4 1 9 2 1 2 1 35 17 1 1 1 1 1 4 0 0 3 4 1 1 3 0 1 3
41681 4168 129.55323 227.21106 1 2 1 10 1 1 1 1 35 17 1 1 1 1 1 4 0 0 3 4 1 3 3 0 1 2
41701 4170 74.62077 137.37529 1 4 1 9 2 2 3 2 35 17 1 1 1 1 1 4 0 0 3 3 1 0 3 0 1 2
```

Understanding your Data

- Read Key carefully
  - Variable information
    - race, sex, income, residence, immunization status
  - Helps when coding
- Not all data is useful for analysis
  - PDAT == 0
  - 471 to 277 children

```
Section             1
Variable Name       PDAT
Label               CHILD HAS ADEQUATE PROVIDER DATA
Response
Categories

                              Adequate
                              Provider
                    All         Data       Value  Label

                  21310         21310           1  CHILD HAS ADEQUATE PROVIDER DATA
                   9620             0           2  CHILD DOES NOT HAVE ADEQUATE
                                                   PROVIDER DATA

Type                NUM
Location            0036  - 0036
Width               8
Sequence Order      6
Notes               USE PDAT = 1 TO IDENTIFY CHILDREN WITH ADEQUATE PROVIDER DATA
                   (INCLUDES UNVACCINATED CHILDREN).
```

Code in R

- Goal: Estimate DTP shots
  - Horvitz-Thompson Estimator
- Using simple random sampling method
  - Equal probability
- Sample size of 50 children

```r
N = nrow(imm) # number of rows in imm data in context of data
# it means the number of children # in the imm data set

n = 50 # variable created to represents arbitrary sample size

imm$P_NUMDTP # displays P_NUMDTP variable entries in imm data
# P_NUMDTP represents the number of DTP shots a child has been given so far

sum(imm$P_NUMDTP) # sums all entries in P_NUMDTP meaning it gives the total
# number of shots all the children combined have had

(N/n)*sum(imm$P_NUMDTP) # prior sum multiplied by total respondents and
# inclusion probability

imm$P_NUMDTP[1:50] # displays P_NUMDTP data again but only the first 50 entries from the full set

imm$P_NUMDTP[sample(1:277, 50, replace = FALSE)] # displays P_NUMDTP data but randomly
# selected entries from a sample created of size 50

(N/n)*sum(imm$P_NUMDTP[sample(1:277, 50, replace = FALSE)]) # outputs an estimated number of
# DTP shots given to whole imm population

(1/n)*sum(imm$P_NUMDTP[sample(1:277, 50, replace = FALSE)]) # prints estimation of
# the number of DTP shots per child in Chicago
```

Thank You

file:///Users/mekiaskebede/Desktop/DRPSTATS/DRP-Final-Report-PDF3.pdf