Horvitz-Thompson (HT) and Yates-Grundy(YT) Hurwitz (HH) and Hajek(Ha) ESTIMATOR (Lab 9)

A.Mayuri(2348133)

R-Markdown

HT AND YT Estimators(Lab 9)

A.Mayuri(2348133)

2023-09-26

About The Data:

The result of sample survey on the number of bearing lime trees and the area reported under limes, in each of the 22 villages growing lime in one of the tehsils of Bangalore district, are given below:

Variable of Interest:

Area Under lime(in acres): Gives the area under which the number of trees with limes are recorded.

No. of bearing lime trees: is the head count of the number of lime trees in the region.

Objective:

1. Give estimates for the average number of bearing lime tress using exact variance estimates- Horvitz-Thompson(HT) and Yates-Grundy(yt) and suggest which estimator is more efficient.

2. Estimate total number of bearing lime trees using approximate variance estimates- Hansen- Hurwitz (HH) and Hajek(Ha) estimators along with the 95% confidence interval. Comment on which estimator provides the estimated total number of bearing lime trees with minimum bias and is more efficient among all 4 estimators.

ANALYSIS:

Step1:IMPORT DATASET

```
library(readxl)

lab <- read_excel("C:/Users/mayur/Desktop/Mstat/tri sem 1/R/dataset/lab.xlsx")

View(lab)

attach(lab)
```

Step2: INITIATING THE LIBRARY

```
library(samplingbook)

## Loading required package: pps

## Loading required package: sampling

## Loading required package: survey

## Loading required package: grid

## Loading required package: Matrix

## Loading required package: survival

##

## Attaching package: 'survival'

## The following objects are masked from 'package:sampling':

##

##     cluster, strata
```

```
##
## Attaching package: 'survey'

## The following object is masked from 'package:graphics':
##
##     dotchart

summary(lab)

##  Area Under lime(in acres) No. of bearing lime trees
##  Min.   :  0.290           Min.   :    0.0
##  1st Qu.:  2.038           1st Qu.:  110.8
##  Median :  6.270           Median :  534.5
##  Mean   : 22.596           Mean   : 1465.5
##  3rd Qu.: 38.215           3rd Qu.: 2180.0
##  Max.   :123.360           Max.   :11799.0

#Suppose we want to estimate the average area
#by considering the population as a auxiliary variable
set.seed(123) # to fix the sample
pps= pps.sampling(z=lab$`Area Under lime(in acres)`,n=20,
method='midzuno')
pps

##
## pps.sampling object: Sample with probabilities proportional to size
## Method of Midzuno:
```

```
##
## PPS sample:
##  [1]  1  2  4  5  6  7  8  9 10 11 12 13 14 15 16 18 19 20 21 22
##
## Sample probabilities:
##            [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
##  [1,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [2,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [3,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [4,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [5,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [6,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [7,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [8,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [9,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [10,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [11,] 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473
## [12,] 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580
## [13,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [14,] 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304
## [15,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [16,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [17,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
```

```
## [18,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [19,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [20,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##            [,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
##  [1,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [2,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [3,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [4,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [5,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [6,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [7,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [8,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [9,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [10,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [11,] 0.7102473 0.7102473 0.7102473 0.7102473 0.5847521 0.7102473 0.2632275
## [12,] 0.8692580 0.8692580 0.8692580 0.5847521 0.8692580 0.8692580 0.3689242
## [13,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [14,] 0.4558304 0.4558304 0.4558304 0.2632275 0.3689242 0.4558304 0.4558304
## [15,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [16,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [17,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [18,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [19,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
```

```
## [20,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##           [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
##  [1,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [2,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [3,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [4,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [5,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [6,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [7,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [8,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [9,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [10,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [11,] 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473
## [12,] 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580
## [13,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [14,] 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304
## [15,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [16,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [17,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [18,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [19,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [20,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
```

```
sample= lab[pps$sample,]

sample #selected sample using midzuno method

## # A tibble: 20 × 2
##    `Area Under lime(in acres)` `No. of bearing lime trees`
##                          <dbl>                       <dbl>
##  1                        32.8                        2328
##  2                         7.97                        754
##  3                        15.6                         949
##  4                        42.8                        3091
##  5                        40.0                        1736
##  6                         9.39                        840
##  7                         6.33                        311
##  8                         5.05                          0
##  9                        94                          3044
## 10                        53.7                        2438
## 11                         0.67                        128
## 12                         0.82                        102
## 13                         2.15                         60
## 14                         0.43                          0
## 15                       123.                        11799
## 16                         3                           317
## 17                         4                           190
## 18                         2                           180
```

```
## 19                    6.21                    752
## 20                    45.8                    3091
```

HT and YT Estimate

```
#HT estimate with different methods of variance estimation
## Exact variance estimate-Horvitz-Thompson and Yates-Grundy


N=nrow(lab)
N

## [1] 22

PI=pps$PI
PI #PI is the matrix of second order inclusion probabilities of sample

##            [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
##  [1,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [2,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [3,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [4,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [5,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [6,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [7,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [8,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [9,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [10,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
```

```
## [11,] 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473
## [12,] 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580
## [13,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [14,] 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304
## [15,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [16,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [17,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [18,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [19,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [20,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##            [,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
##  [1,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [2,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [3,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [4,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [5,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [6,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [7,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [8,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##  [9,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [10,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [11,] 0.7102473 0.7102473 0.7102473 0.7102473 0.5847521 0.7102473 0.2632275
## [12,] 0.8692580 0.8692580 0.8692580 0.5847521 0.8692580 0.8692580 0.3689242
```

```
## [13,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [14,] 0.4558304 0.4558304 0.4558304 0.2632275 0.3689242 0.4558304 0.4558304
## [15,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [16,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [17,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [18,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [19,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
## [20,] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304
##           [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
##  [1,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [2,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [3,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [4,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [5,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [6,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [7,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [8,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
##  [9,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [10,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [11,] 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473 0.7102473
## [12,] 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580 0.8692580
## [13,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [14,] 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304 0.4558304
```

```
## [15,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [16,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [17,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [18,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [19,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [20,] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
```

Estimate Using Horvitz Thompson Method

```
#1. Estimation of variance using Horvitz-Thompson
htestimate(y=sample$`Area Under lime(in acres)`, N=N, PI=PI, method='ht')

##
## htestimate object: Estimator for samples with probabilities proportional to size
## Method of Horvitz-Thompson:
##
## Mean estimator: 22.59591
## Standard Error: 0.0212292
```

The mean estimate of area under treest with lime are 22.59591 and wit variance 0.0212292

```
#2. Estimation of variance using Yates and Grundy
htestimate(sample$`Area Under lime(in acres)`, N=N, PI=PI, method='yg')

##
## htestimate object: Estimator for samples with probabilities proportional to size
## Method of Yates and Grundy:
##
```

```
## Mean estimator: 22.59591

## Standard Error: 5.170035e-18
```

the mean estimate using yates and grundy is 22.59591 and the standard error is 5.170035e-18.

which is lesser than that of HT method

```
## Approximate variance estimate-Hansen-Hurwitz (HH) and Hajek(Ha).

#1. Estimation of variance using Hansen-Hurwitz (HH)

pk= pps$pik[pps$sample]

pk#first order inclusion probabilities of sample units

##  [1] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000

##  [8] 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000 0.4558304

## [15] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000

htestimate(sample$`Area Under lime(in acres)`, N=N, pk=pk, method='hh')

##

## htestimate object: Estimator for samples with probabilities proportional to size

## Method of Hansen-Hurwitz (approximate variance):

##

## Mean estimator: 22.59591

## Standard Error: 6.875034

#2. Estimation of variance using Hajek(Ha)

pik=pps$pik

pik #first order inclusion probabilities of length N for
```

```
##  [1] 1.0000000 1.0000000 0.6572438 1.0000000 1.0000000 1.0000000 1.0000000
##  [8] 1.0000000 1.0000000 1.0000000 1.0000000 0.7102473 0.8692580 1.0000000
## [15] 0.4558304 1.0000000 0.3074205 1.0000000 1.0000000 1.0000000 1.0000000
## [22] 1.0000000

#the population elements
htestimate(sample$`Area Under lime(in acres)`, N=N, pk=pk, pik=pik, method='ha')

##
## htestimate object: Estimator for samples with probabilities proportional to size
## Method of Hajek (approximate variance):
##
## Mean estimator: 22.59591
## Standard Error: 4.677222e-18

# Calculate confidence interval based on normal distribution
#for number of cases i.e. estimated total number of influenza cases
est.ht= htestimate(sample$`Area Under lime(in acres)`, N=N, PI=PI, method='ht')
est.ht

##
## htestimate object: Estimator for samples with probabilities proportional to size
## Method of Horvitz-Thompson:
##
## Mean estimator: 22.59591
## Standard Error: 0.0212292
```

```
est.ht$mean*N #estimated total number of trees

## [1] 497.11
```

Thus the estimated total number of trees is 497.11

```
lower = est.ht$mean*N - qnorm(0.975)*N*est.ht$se
upper = est.ht$mean*N + qnorm(0.975)*N*est.ht$se
c(lower,upper)

## [1] 496.1946 498.0254
```

The confidence interval is 496.1946 498.0254

```
# true number of area under cultivation
sum(lab$`Area Under lime(in acres)`)

## [1] 497.11
```

The population total is equal to that of the estimated total of the number of trees

Conclusion :

HT: Mean estimator: 22.59591 Standard Error: 0.0212292

YG:

Mean estimator: 22.59591 Standard Error: 5.170035e-18

HH:

Mean estimator: 22.59591 Standard Error: 6.875034

HA: Mean estimator: 22.59591 Standard Error: 4.677222e-18

Thus HA Method of Hajek (approximate variance) is the best estimator of the four.