A Beginner's Resource to Capture-Recapture Estimation

Kathryn McGinnis, Prof. James Broda

5/18/2021

An Alternative Estimation Method

Increased rates of homelessness, especially among youth, are a growing concern for American cities. Improper housing, unsafe family situations, and addiction are driving more people each year to be temporarily unhoused. While the need for effective housing policy is great, governments and community agencies must also have proper evaluation programs to monitor the impact of these policies.

The Department of Housing and Urban Development (HUD) conducts an annual Point-In-Time (PIT) survey to estimate the number of homeless youth in a given area. The data is an important metric for evaluating a community's need, but it is a limited picture of people experiencing homelessness at any point in time. The surveyor only counts the number of people *observed*

To estimate a closed population Capture-Recapture experiment, a researcher must record when a member of the population was observed and whether or not they have been seen before throughout the data collection period. The actual data can take many forms, including cross referencing unique lists, recording identifying characteristics of population members, or asking people to recall images they have seen in the past.

Example Data Collection: Allegheny County, PA Lesbian Population

This data was taken from a 2002 study completed by Aaron, Chang, Markovic, and LaPorte to estimate the number of lesbians in a Pennsylvania county. The researchers analyzed four mailing lists from LGBTQ community centers and organizations throughout the county, noting the number of women that appear on multiple lists. In this design, each list will serve as a separate observation period.

The data for this experiment are available to the public and preloaded into the Rcapture software; however,

is sighted throughout the observation period.

First, aggregate each individual list of names to create a master list.

```
#Label each list
a<-df[,1]
b<-df[,2]
c<- df[,3]

#Create one master list
total<-c(a,b,c)
Total<-total[total !=""]
Total
```

```
##  [1] "G. Washington" "B. Franklin"   "A. Hamilton"   "T. Jefferson"
##  [5] "P. Revere"     "B. Arnold"     "S. Adams"      "N. Greene"
##  [9] "P. Henry"      "N. Hale"       "C. Attucks"    "J. Hancock"
## [13] "H. Knox"       "D. Morgan"     "G. Washington" "T. Jefferson"
## [17] "B. Arnold"     "S. Adams"      "J. Dickinson"  "J. Trumbull"
## [21] "J. Brant"      "H. Gates"      "J. Sullivan"   "B. Lincoln"
## [25] "B. Ross"       "B. Tarleton"   "F. Marion"     "M. Hays"
## [29] "S. Adams"      "N. Greene"     "J. Trumbull"   "G. Washington"
## [33] "J. Hancock"
```

```
#Notice that this list contains repeated names
#Some people were sighted multiple times during the collection period.
```

Next, identify the number of unique names that appear on each list.

```
#Identify the unique names in the master list
Unique<-unique(Total)
Unique
```

```
##  [1] "G. Washington" "B. Franklin"   "A. Hamilton"   "T. Jefferson"
##  [5] "P. Revere"     "B. Arnold"     "S. Adams"      "N. Greene"
##  [9] "P. Henry"      "N. Hale"       "C. Attucks"    "J. Hancock"
## [13] "H. Knox"       "D. Morgan"     "J. Dickinson"  "J. Trumbull"
## [17] "J. Brant"      "H. Gates"      "J. Sullivan"   "B. Lincoln"
## [21] "B. Ross"       "B. Tarleton"   "F. Marion"     "M. Hays"
```

```
#24 unique names appear on the three lists
```

Now that the program has identified the name of each unique individual, it can search each list for a specific name. If a name appears on a given list, the program will show a "1". If the name is not on the list, the program will show a "0".

```
#Create an empty matrix to store the capture histories

History<-NULL
x1<-rep(0, length(Unique))
x2<-rep(0, length(Unique))
x3<- rep(0, length(Unique))
History<- cbind(x1, x2, x3)


#Search each list in the dataset to see if it contains the unique name
#Example shown for first five unique names
```

```r
x<-NULL
for (k in 1:3){x[k]<- sum(df[,k]== G. Washington )}
n1<-
```

this paper uses a capture history format very similar to the one above.

```
#Examine the capture history matrix for the lesbian data set
lesbian
```

```
##        A B C D freq
##  [1,] 1 1 1 1   23
##  [2,] 1 1 1 0   27
##  [3,] 1 1 0 1   48
##  [4,] 1 1 0 0  104
##  [5,] 1 0 1 1   19
##  [6,] 1 0 1 0   44
##  [7,] 1 0 0 1  143
##  [8,] 1 0 0 0  589
##  [9,] 0 1 1 1   20
## [10,] 0 1 1 0   64
## [11,] 0 1 0 1   53
```

Exploratory Heterogeneity Graph

fi: number of units captured i times

log(fi)

6.0

3.5

1.0   1.5   2.0   2.5   3.0   3.5   4.0

i: number of captures

ui: number of units captured for the first time on occasion i

log(ui)

5.5

1.0   1.5   2.0   2.5   3.0   3.5   4.0